DOCUMENT RESUME

ED 435 639 TM 030 239

TITLE	Tool Kit: Evaluating the Development and Implementation of Standards.
INSTITUTION	Council of Chief State School Officers, Washington, DC.
ISBN	ISBN-1-884037-46-1
PUB DATE	1997-00-00
NOTE	306p.
AVAILABLE FROM	Council of Chief State School Officers, One Massachusetts Avenue, NW, Suite 700, Washington, DC 20001-1431 ($15). Tel: 202-336-7016; Fax: 202-408-8072.; Web site: <http://www.ccsso.org>.
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement; Academic Standards; Criteria; Elementary Secondary Education; *Evaluation Methods; Models; Program Development; Program Implementation

ABSTRACT

The contents of this "Tool Kit" have been developed as resources to support state efforts to develop and implement rigorous standards for student achievement. The components have been planned as tools for planning and evaluating the development and implementation of content and performance standards. The following components are included: (1) models for implementation of standards from content standards to classroom instruction and prototypes in science, mathematics, reading, and writing; (2) a set of analyses of the content and format of state standards in six subject areas; (3) a set of consensus criteria for evaluating standards and the sources from which they were drawn; (4) resources for policy makers on policy related to standards; and (5) resources for policy makers and the public on assessment for determining achievement. The last category includes discussions of evaluating expectations and assessments, the progress of Title I programs, comprehensive assessment systems, and large scale assessments. Contains references in each section and a 17-item bibliography. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 639

Tool Kit

Evaluating the Development and Implementation of Standards

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

P. Delo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

TM030239

Walker

There are **164** publications.

Publication Title	Publication Date	Hardcopy Price
Directory of Education Agencies	1999	$25.00
Improving Mathematics Education Using Results from NAEP and TIMSS	1999	$8.00
First in the World through Federal Education Policy and Funding	1999	FREE
State Education Indicators with a Focus on Title I, 1998	1999	$20.00
Council Quarterly, June 1999	1999	N/A
The Council 1999	1999	FREE
Gaining Ground, January 1999	1999	N/A
Gaining Ground, March 1999	1999	N/A
Gaining Ground, April 1999	1999	N/A
Gaining Ground, May 1999	1999	N/A
Gaining Ground, June 1999	1999	N/A
Gaining Ground, July 1999	1999	N/A
Ensuring Accuracy in Testing for LEP Students: A Practical Guide for Assessment Development (Preview)	1999	N/A
Guide to Scoring LEP Student Responses to Open-Ended Science Items - SCASS LEP Consortium Project	1999	$10.00
Teenage Pregnancy: State Education Agencies and Prevention	1999	FREE
State Education Accountability Reports and Indicator Reports: Status of Reports Across the States 1999	1999	$5.00
Directory of Coordinated School Health Program Staff 1999-2000	1999	FREE
State Indicators of Science and Mathematics Education 1999 - State Trends and New Indicators from the 1997-98 School Year	1999	N/A
Standards Reviews and Benchmarking Services	1998	FREE
Sustaining School-to-Career Systems for Urban Youth: A Consortium of State Urban Partnerships	1998	$-1
The Exchange Newsletter, Winter 1998	1998	FREE
The Exchange Newsletter, Spring 1998	1998	FREE
Incorporating Health-Related Indicators in Education Accountability Systems	1998	$1
Tool Kit: Evaluating the Development and Implementation of Standards	1998	$15.00

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS (CCSSO) is a nationwide, nonprofit organization composed of the public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five extra-state jurisdictions. CCSSO seeks its members' consensus on major educational issues and expresses their view to civic and professional organizations, federal agencies, Congress, and the public. Through its structure of standing and special committees, CCSSO responds to a broad range of concerns about education and provides leadership on major education issues.

Because CCSSO represents each state's chief education administrator, it has access to the educational and governmental establishment in each state and to the national influence that accompanies this unique position. CCSSO forms coalitions with many other education organizations and is able to provide leadership for a variety of policy concerns that affect elementary and secondary education. Thus, CCSSO members are able to act cooperatively on matters vital to the education of America's young people.

The *Tool Kit* was developed by CCSSO with support from the State Education Improvement Partnership (SEIP). SEIP is a collaboration between CCSSO, The Education Commission of the States (ECS), The National Association of State Boards of Education (NASBE), The National Conference of State Legislatures (NCSL), and The National Governors' Association (NGA) and is funded by a grant from the U.S. Department of Education.

For additional information on the *Tool Kit*, please contact Julia MacMillan at (202) 336-7053 or juliam@ccsso.org. For information pertaining to SEIP, please contact Anuradha Sapru Kohls at (202) 336-7069 or anuk@ccsso.org.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
Wilmer S. Cody (Kentucky), President
Robert E. Bartman (Missouri), President-Elect
Henry R. Marockie (West Virginia), Vice President
Gordon M. Ambach, Executive Director

Copies of this document may be ordered from:

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
Attn: Publications
One Massachusetts Avenue, N.W., Suite 700
Washington, DC 20001-1431
Phone: (202) 336-7016
Fax: (202) 408-8072

ISBN 1-884037-46-1

Copyright, 1997 © by the Council of Chief State School Officers, Washington, DC

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

INTRODUCTION 1

IMPLEMENTATION MODELS AND PROTOTYPES

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

POLICY

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

INTRODUCTION

The contents of the Tool Kit have been developed as resources to support state efforts to develop and implement rigorous standards for student achievement. The contents are ever-changing drafts and revisions and additions will be made over time as states provide feedback on their changing needs.

Why the Tool Kit Was Developed

The Tool Kit was developed because states and districts have indicated a need for assistance in evaluating their efforts to develop and implement content standards. Some states and districts have already conducted evaluations using outside experts. As more states and districts complete the initial phases of work in developing standards, the demand for support and resources is likely to increase. There are several publications that address aspects of evaluating standards and a small number of organizations have provided assistance, but there is no single resource or agency for assisting in the review of standards development and implementation. This resource attempts to streamline efforts to find guidance and enable states and districts to design their own processes for review.

How to Use the Tool Kit

The components have been developed to serve as tools for planning and evaluating the development and implementation of content and performance standards. The various components can be used:

- ✔ by state staff and committees in the development stages of standards,
- ✔ for a formative self evaluation of state efforts completed either by a group of state stakeholders or by an outside panel, or
- ✔ to plan and develop aspects of standards implementation such as curriculum materials or assessment resources.

Because the enterprise of systemic reform is being continuously rethought and refined, the contents are subject to revision and any suggestions for changes or resources that should be added would be appreciated. Please contact Julia MacMillan at (202) 336-7053 or juliam@ccsso.org.

Components

1. Models for implementation of standards from content standards to classroom instruction and prototypes in:
   - ✔ science,
   - ✔ mathematics,
   - ✔ reading, and
   - ✔ writing.

2. A set of analyses of the content and format of state standards in:
   - ✔ science,
   - ✔ mathematics,
   - ✔ reading,
   - ✔ writing,
   - ✔ language arts, and
   - ✔ social studies.

3. A set of consensus criteria for evaluating standards and the sources from which they were drawn.

4. Resources for policy makers on policy related to standards.

5. Resources for policy makers and the public on assessment for determining achievement.
   - ✔ Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education
   - ✔ Implementating the Adequate Yearly Progress of Title I
   - ✔ How Should the Comprehensive Assessment System Be Designed?
   - ✔ Guidelines for the Development and Administration of Performance Assessments in Large Scale Assessment Programs

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

INTRODUCTION

2 IMPLEMENTATION MODELS AND PROTOTYPES

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

POLICY

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

IMPLEMENTATION MODELS AND PROTOTYPES

As states move from the development of content standards to their implementation, all aspects of education may undergo some degree of change. However, certain key products and activities emerge directly from the implementation of the content standards in the classroom.

The paradigm illustrated in the flow chart on page 3 indicates that each content standard is the starting point for the development of performance standards, assessment tasks, scoring guides, benchmarks, and instructional activities. Although the flow chart suggests a linear development, the various pieces influence each other, in actuality. In fact, the implementation of standards regarded from another perspective, that of the classroom teacher, for example, looks different. The flow chart on page 4 illustrates the classroom perspective.

Content standards, developed with input from teachers and other educators, tend to represent much of what is already happening in effective classroom instruction. Sound assessment tasks and scoring guides also should reflect the best of classroom instruction. Certainly, a review of student responses to actual assessment tasks often makes it necessary to revise or rethink performance standards as well as performance tasks. However, the purpose of articulating content standards and performance standards is to provide common focus and alignment among all the aspects of education, and the model suggested here illustrates that focus.

Although the process of implementation is not exactly linear, the connections among these various pieces are essential. It is useful for those developing standards to be aware of this implementation model because it reminds them that stating of the content standards and even the performance standards is only an initial step, albeit a crucial one, in a complex process of implementation.

Content and performance standards should be substantive enough to support the improvement of all the aspects of education, but they cannot be so specific that there is no room for interpretation in the subsequent implementation process. In other words, a content standard that describes one definite assessment task may be overly specific. A set of such standards could make the design of an appropriate assessment almost impossible, as it would be difficult to include enough items to determine reliably whether each of many specific content standards has been met and keep the assessment a reasonable length. On the other hand, if a standard is too broad, almost anything counts for achievement, and there is not enough guidance to develop assessment, much less instructional resources.

In the interest of simplicity, some essential items were omitted from the model presented here. For example, staff development resources and plans, as well as teacher certification policies, are important items that this paradigm has not incorporated in order to produce a focused tool. In the future, these components will be added to the paradigm of the implementation process.

The *Tool Kit* focuses on the state perspective. Some states will develop all the pieces using representative educators from local education agencies. Other states will develop the content and performance standards, but leave the creation of assessments and instructional resources to the local education agencies. Districts are also developing and implementing standards, and it is believed that this model will prove useful to their efforts.

The *Tool Kit* provides prototypes of the aligned pieces of standards implementation. The prototypes were developed by content experts in each area. The variation in their approaches was maintained because it is important to acknowledge the diversity with which standards will be implemented.

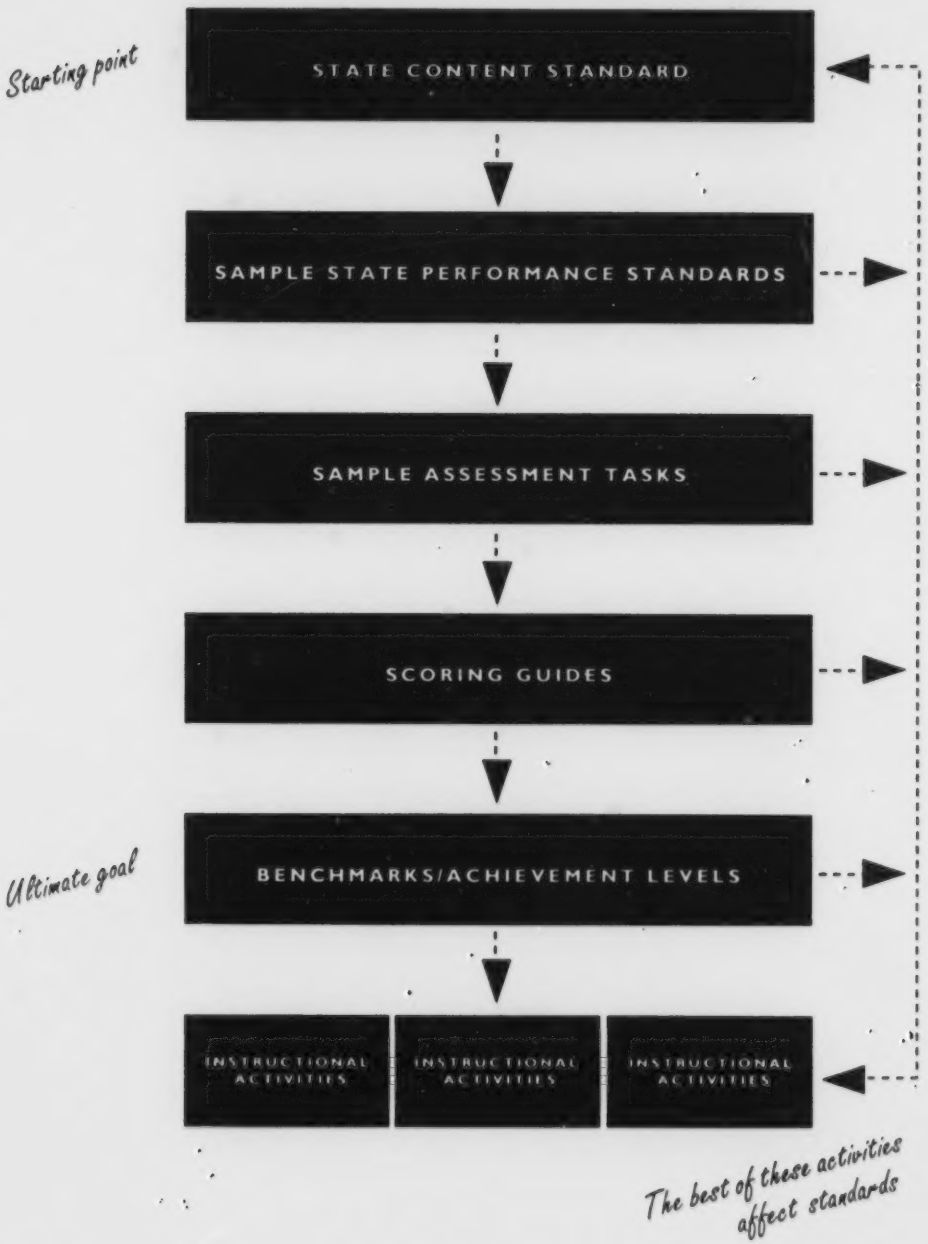
Any additions or suggestions for this component of the *Tool Kit* are welcome.

PAGE 5

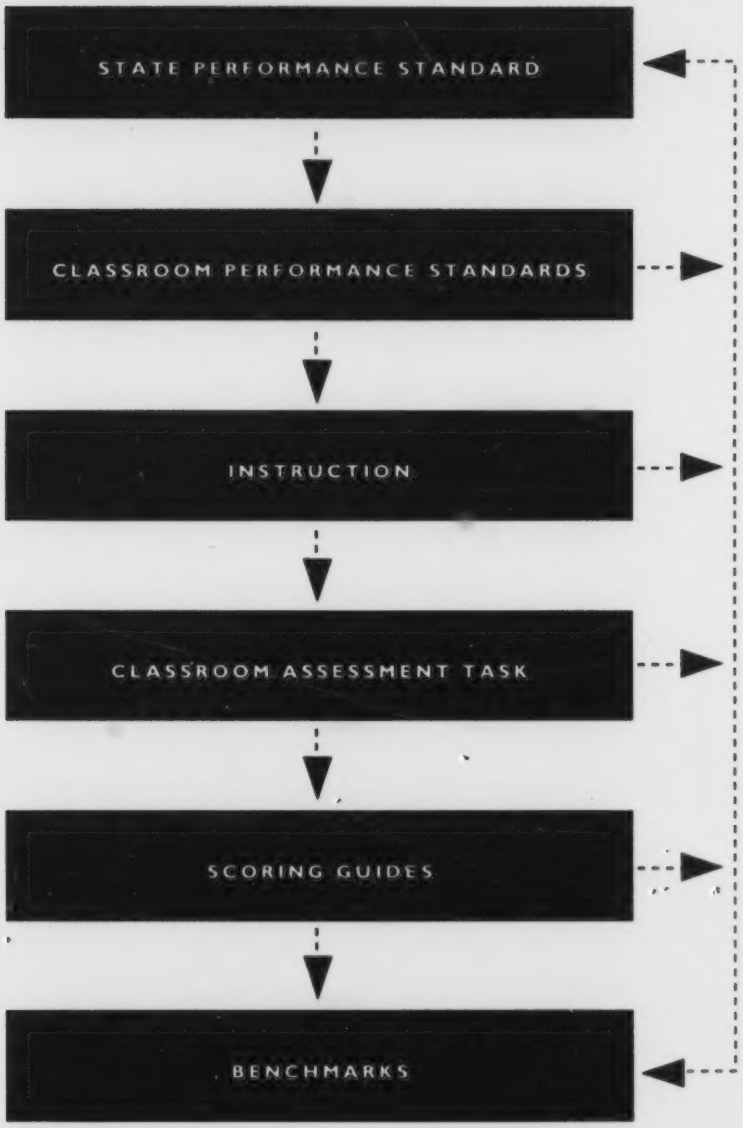
PAGE 29

READING, PAGE 42
WRITING, PAGE 77

State Perspective on Implementation of Standards in the Classroom



Classroom Perspective on Implementation of Standards



National Standard: Energetics

notes

EXAMPLE: LIFE SCIENCE

Organisms require the transformation, transportation, and release of energy. Energy limits the distribution and abundance of organisms in populations and ecosystems. As matter and energy flow through different levels of organization of living systems — cells, organs, organisms, communities — and between living systems and the physical environment, chemical elements are recombined in different ways. Each reconstruction results in storage and dissipation of energy into the environment as heat. Matter and energy are conserved in each change.

The energy for life is primarily derived from the sun. Plants capture energy by absorbing light and using it to form strong chemical bonds between the atoms of carbon-containing molecules. The energy stored in these chemical bonds can be used as sources of energy for life processes.

(Bork, 1995; National Research Council, 1995)

State Standards: The Study of Energy Flow

EXAMPLE: ENERGY FLOW

- ☑ Students examine how living things are tied to one another and to the physical environment by the transfer and transformation of energy.
- ☑ Students use curiosity and open-mindedness to explore, study, test, and modify ideas through scientifically designed and controlled testing.
- ☑ Students explore investigatory pathways that include the invention of ideas, concepts, processes, controlled experiments, the use of previous research and conclusions to advance essential scientific concepts and theories.

(California Department of Education, 1990; National Research Council, 1995)

State Performance Standard

Investigate and explain how cells and organisms react to stimuli and maintain stability (e.g., plant cells in salt solution, bacteria in contact with antibiotics; cell components for photosynthesis, respiration and waste removal; enzyme and hormone actions).

(Preliminary Draft: Illinois Academic Standards, Volume Two, State Goals 11-18, Science, June 1996.)

State Performance Standard

Describe how scientific knowledge, explanations, and technological designs may change with new information over time.

(Preliminary Draft: Illinois Academic Standards, Volume Two, State Goals 11-18, Science, June 1996.)

notes

Classroom Performance Standard

LIFE SCIENCE CONTENT: PHOTOSYNTHESIS

Students demonstrate a thorough understanding of photosynthesis through a detailed discussion of the processes involved in the transformation of light energy to chemical bond energy. Student work effectively communicates the importance of photosynthesis as a source of energy for life processes through all levels of ecosystems.

Classroom Performance Standard

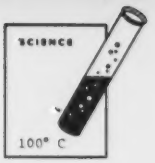
SCIENTIFIC INQUIRY: INVESTIGATING PHOTOSYNTHESIS

Students thoughtfully design and thoroughly implement an investigation to study, verify or research aspects of photosynthesis. They collect and display accurate and appropriate data to illustrate relevant results, clearly identify patterns and trends, and support conclusions. Viable future testing and the significance of photosynthesis to personal, social or ecological issues are effectively communicated.

Classroom Performance Standard

HISTORY AND NATURE OF SCIENCE: STUDY OF THE SCIENTIFIC RESEARCH THAT LED TO CURRENT UNDERSTANDING OF PHOTOSYNTHESIS

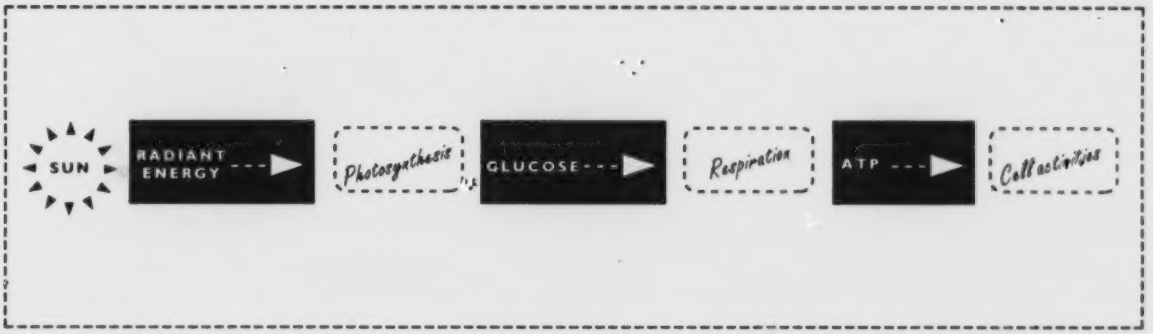
Students demonstrate thorough nonexperimental research techniques, including in-depth review of previously conducted research, data collection, and analysis of historical source documents, to explain the investigatory pathway leading to the current body of knowledge regarding photosynthesis.



Instructional Activities

LIFE SCIENCE CONTENT LESSON

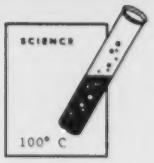
A primary focus of Ms. K's biology class had been to show students the connectedness of the biology curriculum. At the end of the second semester, she realized that she had an opportunity to do just that following a three-week exploration of ecology. Knowing that students had already studied the biochemistry of photosynthesis, she began a 3-day lesson on the cyclical nature of energy and matter transformation through ecosystems.



Ms. K began her lesson with a review of the needs of living things. She assigned groups of students a specific high-order ecosystem consumer and asked them to research and identify the other potential members of the corresponding food chain, and to compare and contrast the needs and characteristics of each member.

On the second day, student groups reported their findings. Each group reported that while all members of their food chain needed energy for cell activities, only the producers in the food chain were able to "make it themselves" by "capturing" the energy of the sun.

On the third day, Ms. K asked the student groups to fill in the details of the above diagram with the products and reactants of the processes they had studied in the previous semester. She also asked them to think about how matter and energy were cycled through the ecosystem described earlier.



Classroom Assessment Task

notes

LIFE SCIENCE CONTENT: INTEGRATED WRITING TASK
CREATIVE WRITING ABOUT PHOTOSYNTHESIS —
WHY IS IT IMPORTANT?

- ✔ We have spent the last 2 weeks learning about tropical rain forests and the plants and animals within them. Photosynthetic plants are the basis for all ecosystems. Your assignment is to creatively explain photosynthesis as a process and the reason that photosynthesis is important to the jaguar found in the rain forest. Your work must be thorough and include detailed scientific information.
- ✔ Examples of creative ideas include:

travel brochure	*video*	*song*
collage	*children's book*	*rap*
short story	*poster*	*whatever (see me)*
photo essay	*poem*	

- ✔ You must use your own words and work independently. If your project is mostly pictures, you need to include a page of explanation.
- ✔ Whatever medium you choose, you must design your project and any explanation so that they are clear and understandable to someone who has not taken a biology class.
- ✔ Your work will be judged on how well you relate scientific facts and concepts to real-world topics and on how much you understand about the processes involved in photosynthesis, and the patterns and trends of energy transfer in the ecosystem.

(California Department of Education, 1993)

Scoring Guide

notes

SCIENTIFIC CONTENT — DEPTH OF CONTENT UNDERSTANDING

LEVEL 4 — Scientific facts, concepts, and ideas are thoroughly and precisely explained through examples, descriptions, analogies, or illustrations. Patterns and trends are identified, discussed and described in terms of issues that are personally, socially or ecologically relevant. Connections among scientific concepts, principles or themes within and across scientific disciplines are identified and described.

LEVEL 3 — Scientific facts, concepts, and ideas are accurately explained through examples, descriptions, analogies, or illustrations. Patterns and trends are identified and related to significant issues. Connections among scientific concepts, principles, or themes within and/or across scientific disciplines are identified.

LEVEL 2 — Scientific facts, concepts, and ideas are explained with minor inaccuracies. Patterns and trends are identified or implied. Connections among scientific concepts, principles, or themes within or across scientific disciplines are suggested or implied.

LEVEL 1 — Scientific facts, concepts, and ideas are discussed with major inaccuracies. Patterns and trends are unclear or inaccurate. Connections among scientific concepts, principles, or themes within or across scientific disciplines are unclear or absent.

(State Collaborative for Assessment and Student Standards [SCASS], 1995; California Department of Education, 1993)

Benchmarks

LIFE SCIENCE CONTENT — STUDENT EXEMPLAR AND COMMENTARY

The work entitled "Leopold Onca" displays strong evidence of the student's broad conceptual understanding of photosynthesis. The student clearly illustrates patterns and trends in the description of the energy transfer throughout the ecosystem at all levels of organization — biochemical, cellular, tissue, and organismal. The work creatively references ecology, chemistry, animal development and adaptation in a meaningful way — highlighting the student's ability to thoughtfully connect many aspects of life science. The student's selection of a creative writing project conveys a general broad and applied understanding, the reflective explanation shows extraordinary depth of precise, accurate scientific knowledge.

Leopold Onca

by Student "L"

Biology 5H

"L" Desktop Publishing

Copyright © 1995 by Student "L"

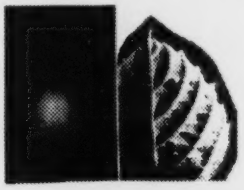
All rights reserved, including the right of reproduction in whole or in part in any form.

Published by: Student "L"

Anytown, USA 90000

Once upon a time (and all the time) a particle of radiant energy traveled from the burning globe of a midsize star called sun to an utterly insignificant little blue-green planet in the uncharted backwaters of the Milky way galaxy. This planet was known (by a large number of ape-descended life forms) as Earth. The particle of energy they recognized as a sunbeam. This particular particle fell on a rather large autotrophic organic substance (better known as a plant) growing in the canopy of a rather beautiful, majestic and almost untouched section of the Amazon rain forest. A new adventure was about to begin for the energy particle. Upon reaching the plant, the energy was trapped by a green pigment (a chlorophyll molecule). It found itself inside what looked like a stack of coins (the grana) that were jammed every where inside a juicy plump bag (the chloroplast) inside the leaf of the plant. But it didn't say with the chlorophyll for long. The chloroplast was full of more than chlorophyll pigments. There were little water and carbon dioxide molecules pushing through the surface constantly. The energy particle was used to rearrange the hydrogen and the oxygen (from 6 waters) and the carbon and oxygen (from 6 carbon dioxides) into a bigger molecule. The energy particle was trapped! It was no longer free to move on its own. It went wherever the bigger molecule went. This new molecule went by the name $C_6H_12O_6$. Millions and millions of molecules of sugar ($C_6H_12O_6$) were put together inside the plant.

$C_6H_{12}O_6$ was a traveling kind of molecule. Soon it was being pulled out of the leaf and into tubes that seemed like a highway system inside the plant. The energy particles, now chemical bond energy, would serve its master, the plant, in many different ways. Sometimes the sugar was stored in plant parts that seemed to be nothing more than bags of sugar water. Sometimes the sugar went to cells where it broke back down into water and carbon dioxide. But the chemical bond energy was put into an even bigger molecule called adenosine triphosphate (ATP). ATP was used wherever it was needed. Sometimes the chemical bond energy was used to make more copies of the biggest molecule of all — the cell's planning molecule — deoxyribonucleic acid (DNA). Sometimes the chemical bond energy was used to make message molecules — ribonucleic acid (RNA). Sometimes it helped stick amino acids together to make proteins for the plant. The possibilities for the use of this little energy particle stagger the imagination. Using all of these chemical reactions, the plant grew and reproduced.



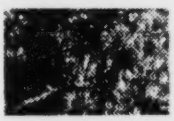
One day, the leaf that the particle of energy had contributed to, was attacked by a larval butterfly (better known as a caterpillar).

The energy was transferred to the caterpillar's digestive system where it was changed yet again in to energy for the caterpillar whom it helped grow and spin a cocoon and change into a butterfly.

This butterfly was flying around advertising for a mate when a tree frog began to watch him. Coming alight on the reassuring surface of a hardwood branch the butterfly began to preen. It was then that the frog struck. Whipping out her long, sticky tongue she stuffed the butterfly into her mouth and swallowed. It was there that the particle of energy became part of the frog.

It contributed to the frog's leaps and jumps until one day the frog leaped right in front of a particularly bloodthirsty herd of peccaries (better known as wild pigs). There the frog was eaten. Although it was unusual, she had been extremely venturesome that day, and had not remembered to look both ways before crossing a clearing. The peccary could do nothing but eat her. It was a particularly weak peccary whose mother had died very soon after his birth. He had been raised by the herd and was rather skinny and timid. His system much appreciated the particle of energy that the frog contributed. However, because peccaries are such supportive group mammals he used the particle of energy to father a litter of baby peccaries. Of the litter there was one that was particularly a favorite of his father, named Wagner.

It was a tragedy to his father when one day Wagner wandered off from the herd and was found by Leopold Onca the jaguar. Leo didn't let Wagner know he was coming but stalked Wagner wandering through the forest. From behind vines he watched, and when Wagner wandered under a convenient tree, Leo sprung from the branch above. Wagner was dead before you could say "ribulose bisphosphate," but the particle of energy was transferred into Leo.



By now the particle of energy was well worn out from being in fauna (animals), it looked forward to cycling up through the flora, the roots of some tree, after Leo had died. Unfortunately this was not to be. You can see the head of the jaguar that the particle helped make on the wall, as a poacher's hunting trophy.

BEST COPY AVAILABLE

STUDENT REFLECTION: LEOPOLD ONCA

In my Science Portfolio I have included a sample creative writing called "Leopold Onca". This work was for a project that I did at the end of a two week biology unit about the rainforest and its plant and animals. The assignment our teacher gave was this — "explain photosynthesis as a process and why photosynthesis is important to the jaguar found in the rainforest."

The scientific information that relates photosynthesis to the jaguar is organized into a creative story that starts with a particle of energy from the sun. That's a good place to start the story because the sun IS the source of all energy on earth. This way of organizing the information is a great way to introduce a lot of science in a little space and in a way that will not scare away a young kid. This book would then make it easy for younger students to be introduced into just how complex the rainforest really is. It would also be enough, I think, for an older student who does not know a lot of science, to read this story and then wonder how these things happen and then go read some more in a textbook or encyclopedia.

The story traces the path of an energy particle as it moves through space to a plant and gets made into chemical energy that helps the plant grow. The plant is then eaten by a caterpillar which changes into a butterfly and is eaten by a frog which in turn is eaten by a peccary which is eaten by a jaguar.

its powerful hind leg muscles to leap and jump. Peccaries travel in herds and show behaviors that help out the population. Jaguar behaviors such as stalking, camouflage, and tree climbing are used to capture their prey.

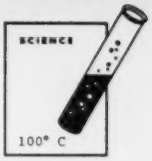
My description of what happens to the energy particle inside a cell shows my understanding of biology of cells and of biochemistry. I described the chemical details while the energy particle was inside a plant cell only. I did this because in this story only plants can do the photosynthesis step of converting carbon dioxide, water and energy into sugars. But the other steps that I describe, sugars being broken down and the energy being captured by ATP molecules, and the ATP molecules being used to basically run the cell — all of these steps happen in every kind of cell.

The creative story shows how important it is that of energy cycle in our universe from sun to planet autotrophs (plants) to heterotrophs (the butterfly, frog, peccary, jaguar) to decomposers and back to autotrophs. It also describes how this process can get interrupted when humans cell the need to hunt animals and display dead animal parts as trophies.

At the end, the energy particle wishes that it could continue in the "natural" cycle of decay and decomposition and be returned into the soil. BUT, the jaguar is killed, its head is mounted as a trophy and the trapped energy particle will not get that chance. At each step of the story I always relate how the energy particle is being used by the particular living organism to carry out some function necessary for its survival. This summary of the story shows both the energy cycling and the food chain.

My vivid descriptions of what each organism does is a way of showing insect metamorphosis, adaptations, and animal behavior. Insect metamorphosis is shown in the story when the caterpillar uses the energy to grow and spin a cocoon and change into a butterfly. Adaptations are shown for both plants and animals. Plants have special structures (grana) filled with the pigment (chlorophyll) that captures the sun's energy. The grana are inside special cells (chloroplasts) that make food for the plant using energy plus water and carbon dioxide (photosynthesis). Some plants (vascular plants) also have a systems of transportation tubes — xylem tissue for transporting water and minerals up from the roots and phloem tissue for carrying manufactured products to all parts of the plant. A butterfly's colorful wings are an adaptation to advertise for mates. The frog has a long, sticky, coated tongue to zap and capture food. The frog moves (but not always fast enough) by using

BEST COPY AVAILABLE



Instructional Activities

NATURE OF SCIENCE LESSON: PHOTOSYNTHESIS

Ms. M believes that an understanding of the history of scientific ideas enriches the understanding of the nature of scientific inquiry. She wants her students to understand how ideas in science develop and change under the influence of the values, ideas, and resources prevalent in society at any given time. She uses a historical approach to introduce an important concept in life science.

Ms. M was beginning the second round of planning for the high-school biology class. Students came to the class with some knowledge and understanding about green plants, but they still had many questions. As a way to help students focus some of their questions, and to highlight the interdependence of science and civilization, she planned to begin the unit with a lecture on photosynthesis. Lecturing was something she seldom did. However, the purpose here was not to lay out the details of the photosynthetic process, but to illustrate how the scientific community's knowledge of photosynthesis had changed over time.

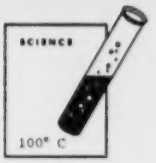
She began the lecture by putting a transparency on the overhead projector of that detailed diagram of photosynthesis (see page 7). She wanted the students to observe the complexity of the current knowledge about photosynthesis, and this diagram was a useful introduction to her lecture. She asked the students how long the scientific community had known about these many complex reactions, why this knowledge was important, how they had come to know so much, and if there was still more detail to be described.

Next, she asked the students to tell her what they already knew about photosynthesis. She expected that most would recall that carbon dioxide, water, sugar, oxygen and sunlight were important and many would recall growing plants in dark cupboards and under boxes in middle school. The next two questions would have to be worded carefully: First, why is photosynthesis so important or, put another way, what is the fundamental question that photosynthesis answers? Second, how long have scientists known about photosynthesis?

With this introduction, she lectured about the seventeenth century experiment of van Helmont, his tree, and his conclusion that the weight of the plants came from water. Ms. M paused. "Was van Helmont wrong?" she asked the students. She expected them to have difficulty conceiving that van Helmont had conducted an experiment, which they knew was essential to science, but that he had not obtained the answer they knew was correct. Ms. M helped them analyze the experiment and the conclusions that could legitimately be drawn from it. She then introduced more of the context of van Helmont's investigation: the prevailing belief about plants as a combination of fire and earth, and the design of van Helmont's study to refute this belief.

She then continued her historical lecture using similar details from several other episodes. She described how chemists had learned to collect gases from chemical reactions; she explained how both the advancing scientific technology and the experimental results of Priestley, Ingenhousz, de Sassure and Sutton created an interdependent investigatory pathway leading to our current understanding. Ms. M stopped her history here. Students reviewed how various factors had shaped the development of early knowledge about photosynthesis. Ms. M recorded and organized their views on the chalkboard. From this, they developed a set of questions for continuing the history on their own.

Ms. M had collected a number of textbooks from different periods in the century. She introduced them as a resource for sketching the changing status of knowledge about photosynthesis. She asked the students to work in groups of five (National Research Council, 1995).



Classroom Assessment Task

notes

Each group of five students prepares a brief presentation on the advancement of ideas regarding photosynthesis during a particular historical period. After 2 days to gather information, the groups share the results of their research, and together, they identify or infer what discoveries have been made in each period. Then, using the questions that they formulated during the earlier class discussion, the students return to their groups to determine how each discovery occurred. They identify factors such as new technologies that were relevant to conducting investigations, the sources of funding for various research projects, the personal interests of researchers, occasions of luck or chance, and the theories that guided the patterns that they uncovered, and how they reached their conclusions.

On the next day, group members are given 15 minutes to discuss the following questions as they pertain to a historical figure of their choosing, based on who contributed to the body of knowledge regarding photosynthesis that was shared in the previous class period.

- ✔ What factors — personal, technological, cultural, and/or scientific — led this person to the investigation?
- ✔ How did the results of the investigation influence the investigator, fellow investigators, and society more broadly?
- ✔ What element of this episode seems most characteristic or most revealing about the process of science? Why?

After the 15-minute discussion, each individual student is required to independently respond to the following questions:

- ✔ How was the investigation designed, and why was it designed as it was?
- ✔ What element of this episode seems most characteristic or most revealing about the process of science? Why?

(National Research Council, 1995)

Scoring Guide

notes

HISTORY AND NATURE OF SCIENCE SCORING GUIDE

LEVEL 4 — A question, problem, or historical study is clearly identified. Context and background are provided to explain why the question, problem, or study is relevant. Legitimate and reliable multiple sources of information have been located and used. Insightful applications to personal and societal issues are identified and described in detail. Background information provides a clear context for interpretation. Information and/or data are related to the question, problem, or issue. The consequences of research are identified and discussed. Scientific terminology is used correctly. A rationale is provided for the conclusion. Research outcomes are related to analysis, evaluation, and synthesis of data and/or information. Written expression is excellent and, when appropriate, includes accurate diagrams, charts, tables, graphs, or other media that enhance the communication of specific ideas.

LEVEL 3 — A question, problem, or historical study is identified. Context and background are provided to explain why the question, problem, or study is relevant and/or important to self, society, technology, and/or nature. Multiple sources of information have been identified and used. Applications to personal or societal issues are identified. Background information provides context interpretation. Consequences and alternatives are identified. When appropriate, tables, graphs, and/or charts are used to organize and present the information and/or data collected. Information and/or data are related to the question, problem, or issue. The consequences of research are identified and discussed. Scientific terminology is used correctly. When appropriate, a conclusion is reached and a rationale provided for it. The conclusion or decision is related to analysis, evaluation, and synthesis of data and/or information. Written expression is good and, when appropriate, includes accurate diagrams, charts, tables, graphs or other media that enhance the communication of scientific ideas.

LEVEL 2 — A question, problem, or historical study is clearly identified. Background as to the relevance of the question, problem, or study is limited. Legitimate or reliable multiple sources of information may not have been located and used. Applications to personal or societal issues may be suggested or implied. Background information provides some context for interpretation. Consequences and alternatives are suggested or implied. Information and/or data are not presented clear-

notes

ly or in a well organized format. Tone may be disjointed or encyclopedic. Information and/or data are partially related to the question, problem, or study. Some scientific terminology is used correctly. When appropriate, a conclusion or decision may be reached. That conclusion or decision may be related to analysis of data and/or information. Student communicates a personal opinion OR misrepresents or selectively uses sources of information to support that personal opinion. Emphasis may be on single observations, events, or point of view. Errors or omissions in written expression may lead to incomplete communication.

LEVEL 1 — A question, problem or historical study may not be identified. Only one source of information may have been located and used. Applications to personal or societal issues are unclear or absent. Background information provides minimal context for interpretation. Consequences and alternatives are unclear or absent. Information and/or data are unclear or presented as a list or in an encyclopedic format. Ideas may not be summarized. Information and/or data may not be related to the question, problem, or study. Some scientific terminology may be used. When appropriate, a conclusion or decision may be reached, but it is not supported logically or scientifically. Student communicates a personal opinion. Emphasis is on single observations, events, or point of view. Written expression is flawed and limits effective communication.

(SCASS, 1996; California Department of Education, 1993)

Benchmark

notes

NATURE OF SCIENCE — STUDENT EXEMPLAR AND COMMENTARY

Select the investigative or experimental work of one scientist who contributed to the current body of knowledge regarding photosynthesis and answer the following.

A How was the work of this scientist designed and conducted, and why was it designed as it was?

After the work of Priestley, who discovered the "Earth's ventilation system," was proved "unrepeatable" by other scientists, Jan Ingenhousz discovered that it wasn't plants but other things that turned bad air into good air and thought that different elements were made in the process. He added light to his experiments and because chemistry had be discovered called the bad air carbon dioxide and the good air oxygen.

B What aspect of this episode seems the most characteristic or revealing about the process of science? Why?

I think it was important that Ingenhousz had several steps in his experiment and found out about carbon dioxide. He added one new thing at a time when he changed the experiment each time and used chemistry to describe what happened.

COMMENTARY

The student includes a minimal degree of detail in answering both parts A and B. While she references the work of Priestley in part A, she does not describe it or provide an appropriate background for the experiments of Ingenhousz. Additionally, the specific aspects of experimental design are not described or related to the conclusions. Much information about the work is omitted, and the reader is left to guess how the work contributed to the sequence of events that led to our current understanding. The work illustrates lower performances (level 1-2 work), as it provides little evidence of any understanding that the nature of science is distinct in the use of empirical standards, logical arguments, skepticism, appropriate methods, and truthful reporting.

27

Instructional Activities

notes

LIFE SCIENCE INQUIRY LESSON

Mr. Q strongly believes that when students are asked to 1) question, 2) hypothesize, 3) design and conduct experiments, 4) collect and analyze data, 5) use evidence to support conclusions, and 6) apply those conclusions to life, their science education is be more meaningful.

Mr. Q began his lesson by quoting the work of Joseph Priestly who, in 1772 wrote:

> *I have been so happy as by accident to have hit upon a method of restoring air which has been injured by the burning of a candle and to have discolored at least one of the restoratives which nature employs for this purpose.*

He demonstrated several of the experiments of Priestly and others that led to the current body of knowledge regarding photosynthesis. His objectives were threefold:

1. Review the biochemical processes of photosynthesis and its significance to most living things.
2. Illustrate the importance of confirming scientific research and comparing and contrasting confirmation investigations with self-designed experimentation with unknown results.
3. Demonstrate appropriate scientific techniques that the students can employ to continue to study the process of photosynthesis.

On the next day, the class discussed the demonstrations and brainstormed additional questions that they would like to study and potential methods that they could employ to find answers to those questions.

Classroom Assessment Task

notes

STUDENT-DESIGNED LAB INVESTIGATION

☑ Question A. How does light affect photosynthesis?

Leaves contain pigments that absorb energy from the sun in photosynthesis. Starch is one of the compounds stored in the leaves. What happens when sunlight falls on a leaf? Is all of the light used by the leaf? What effect does the amount or wavelengths of light available have on the amount of product (starch, oxygen, water) produced?

☑ Question B. Is chlorophyll necessary for photosynthesis?

The silver-leaf geranium has green and white leaves (variegated). These are leaves in which pigments appear to be only in certain areas. But are these areas really without pigment? If they were, each leaf would make an excellent organ with which to test whether chlorophyll is necessary for photosynthesis.

☑ Design a detailed hypothesis and experimental procedure to test Question A.

OR

☑ Design a detailed hypothesis and experimental procedure to test Question B.

You must conduct 1 or 2 with materials that you have used in lab or that I approve in advance. I will provide any materials already present in lab and provide suggestions for your design. You are responsible for maintaining the experiment on your own.

☑ Grades are not dependent on conclusions supporting your hypothesis, only that appropriate scientific procedures are observed, accurate data collected, and reasonable discussion of the methods, results and conclusions are included.

☑ You should design an experiment that is different from the experiments we have studied, one for which you do not already know what will be the outcome.

Scoring Guide

notes

SCIENTIFIC INQUIRY

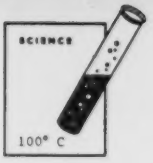
LEVEL 4 — The question, problem, or issue is clearly stated, supported, and logically derived. Research methods generate valid and reliable data that are clearly related to the hypothesis, questions, problem, or issues. Insightful conclusions are clearly stated and are supported by the data. A comprehensive rationale for conclusions includes a recognition of assumptions, multiple interpretations, and an explanation of relationships to scientific principles. Critical evaluation of conclusions is comprehensive and identifies experimental weaknesses, exceptions, contradictory data, consequences, and trade-offs. Logical next-step strategies alter the experimental design so as to yield more conclusive data, comprehensive results, and significant conclusions.

LEVEL 3 — The question, problem, or issue is identified. Hypotheses are stated and supported. Research methods generate reasonable data that are related to the hypothesis, questions, problem, or issue. Appropriate conclusions are clearly stated and are consistent with the data. The rationale for the conclusions includes a recognition of some assumptions, alternative interpretations, and identifications of relationships to scientific principles. Critical evaluation of conclusions identifies some aspects of experimental weakness, exceptions, contradictory data, consequences or trade-offs. Logical next-step strategies alter the experimental design so as to yield more conclusive data, comprehensive results or significant conclusions.

LEVEL 2 — The question, problem, or issue is implied rather than stated. Research methods generate somewhat reasonable data that are somewhat related to the hypothesis, questions, problem, or issue. Conclusions are related to the data. The rationale for the conclusions offers a limited recognition of assumptions, alternative interpretations or relationships to scientific principles. Little justification is used to evaluate or support conclusions. Next-step strategies are proposed and are related to current experimental design.

LEVEL 1 — The question, problem, or issue is unclear or unidentified. Hypotheses are absent or unclear. Research methods generate invalid, unreliable, or unrelated data. Conclusions, if present, are unclear, unsupported, or inconsistent with the data. No rationale is provided for conclusions, decisions, or evaluation of information. Next-step strategies are absent, unclear or unrelated to the current work.

(SCASS, 1995; California Department of Education, 1993)



Benchmark

SCIENTIFIC INQUIRY — STUDENT EXEMPLAR AND COMMENTARY

PHOTOSYNTHESIS LAB

Introduction:

Photosynthesis is an important process that is carried out within the chloroplast of plants. This reaction uses 6 CO_2 and the electrons from six water molecules (6 H_2O) to make glucose ($C_6H_{12}O_6$) and free oxygen (O_2), which will be explained in the conclusion. This reaction is powered by light, as the name implies (photo means light). The purpose of this lab is to discover how different wavelengths of visible light; red, orange, blue, and so on affect the process of photosynthesis. My hypothesis is that the water plant elodea will produce the lowest rate of photosynthesis when exposed to the green light, the highest with full spectrum white light, (because we see that green light is reflected by the plant and not absorbed by the plant) with the other wavelengths somewhere in between.

Procedures:

The elodea plants will be placed in a solution of sodium carbonate in a test tube. An identical test tube of sodium carbonate without any plant material will be used to demonstrate that the sodium carbonate alone produces no reaction when exposed to the various wavelengths of light.

1. Place a 5 cm length of elodea in a test tube, then carefully add 10 ml of sodium carbonate, making sure that no bubbles form around the plant. If bubbles do appear, gently flick the tube until they disappear. Place the test tube in the test tube rack.
2. Prepare a second tube of 10 ml of sodium carbonate, leaving the plant out, and place it in the test tube rack next to the first tube.
3. Shine a full spectrum white light on the two tubes for 3 minutes, and count the number of bubbles which float to the surface. Record your results in the data table below.
4. Repeat the experiment, using colored cellophane over the white light source until all of the colors of cellophane have been tested and recorded.

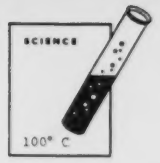
Results:

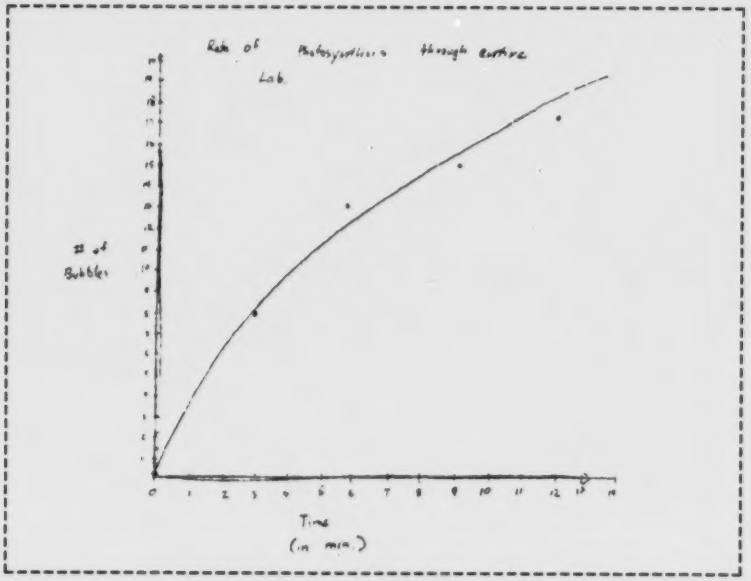
During the test we saw bubbles form on the elodea leaves and counted only those bubbles that actually broke free from the leaves and floated to the surface of the sodium carbonate solution. The bubbles were produced the fastest when the white light was shining on the plant. No bubbles were produced in any of the plain sodium carbonate tubes.

Color of Light	# of bubbles over 3 min.
White	8
Red	5
Green	2
Yellow	3
Blue	2

Rate of Photosynthesis

BEST COPY AVAILABLE





Conclusion:

We expected the green light to produce the least or no bubbles from the stem. Our hypothesis was partially correct. The tubes exposed to the green and blue light each gave off only two bubble.. The tube exposed to the yellow light gave off 3 bubbles, red light gave off 5, and the white light gave off 8 bubbles.

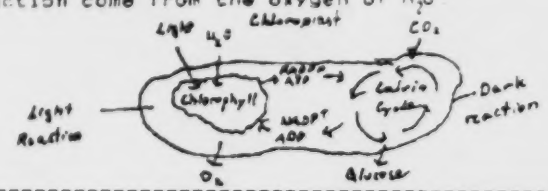
The concept of this lab is photosynthesis. Photosynthesis is the process in which carbon dioxide and water is turned to simple sugars and oxygen. In this reaction, chlorophyll (the catalyst) and light (the energy source) must be present.

$$6\,CO_2 + 12\,H_2O \rightarrow C_6H_{12}O_6 + 6\,O_2 + 6\,H_2O$$

(Carbon Dioxide) + (Water) → (Glucose) + (Oxygen) + (Water)

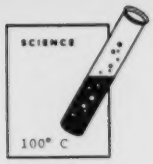
In this lab, we soaked the plant stem in water and sodium carbonate. Sodium carbonate provided carbon dioxide within the water, giving the plant the materials needed for photosynthesis to occur. Photosynthesis occurs in two basic reactions, light and dark reactions. The light reaction occurs first, oxidizing water to oxygen, and the dark reaction combines the hydrogen with carbon dioxide to form simple carbohydrates. Oxygen is given off as a waste product, which were what the bubbles we saw coming out of the plant stem were. The more bubbles, the more oxygen is produced from the plant and in turn, the more oxygen, the faster the rate of photosynthesis of the plant because ~~one~~ H_2O is being split to O_2 in the light reaction.

Within, the chloroplasts of the plant, there ~~are~~ chlorophyll, which is capable of absorbing light energy and exciting water molecules, oxidizing H_2O to O_2. As water is being oxidized into oxygen, energy rich ATP and NADPH are produced. Then, in the dark reaction, this energy rich ATP and NADPH are ~~reduced~~ used in the Calvin cycle to convert CO_2 to carbohydrates. The electrons of this reaction come from the oxygen of H_2O.



NADPH and ATP are high energy after Calvin Cycle
NADP+ and ADP are low energy

BEST COPY AVAILABLE



For photosynthesis to be most efficient, two wavelengths of light must be present, one wavelength greater than 650nm (red) and one wavelength less than 650nm.. This is why the white light had the fastest rate of photosynthesis. White light contains the whole visible color spectrum. In our lab, it seems that light with lower wavelengths don't supply energy as well as light with higher wavelengths since green, yellow, and blue light all had a slower rate of photosynthesis. We know this because our plant produced the least oxygen under these light conditions.

All life on earth is dependant on photosynthesis, directly or indirectly. Plants are the bottom of the food chain. They make food for themselves and are food for higher life forms. If all photosynthesis-capable life forms disappear, animals would slowly, one by one would die of starvation. Also plants, during photosynthesis, produce oxygen for oxygen breathing life forms. Animals use oxygen and create carbon dioxide as waste while plants complete the cycle by using carbon dioxide and creating oxygen as waste. Without plants, we would have no oxygen to breath.

Our conclusion is that different wavelengths of light does effect the rate of photosynthesis: generally, the higher the wavelength, the faster the rate of photosynthesis and the lower the wavelength, the slower the rate of photosynthesis. Our background information states that violet would also be an effective wavelength for photosynthesis. Unfortunately, we did not get a chance to test this color.

If there are any errors in conducting the experiment, it may have been in the experimental design. We could have weighed the plants as an additional control. We could have let the tubes stand in the light longer than 3 minutes, but I don't think either of these changes would have significantly changed the reesults of the experiment.

FURTHER TOPICS OF STUDY

*Does the intensity of the light effect the rate of photosynthesis?

BEST COPY AVAILABLE

GSE SELF-REFLECTION SHEET

1. *Thoroughly* explain the scientific concept you are investigating in this entry. Give specific examples that show how this concept relates to your Problem-solving Investigation.

Every time we take in a breath of air, we breath in oxygen. This oxygen is nc infinite. It comes from plants carrying out photosynthesis, a complicated and beautiful process in which water and carbon dioxide is turned to oxygen and glucose. We breath in the oxygen, a waste product given off from the plants, while the plant uses the glucose. Normally, what powers this process is light from the sun. The purpose of this investigation is to explore how different wavelengths of light, or different power sources, effect the rate of photosynthesis. We investigated this by measuring the output of oxygen under different light sources. The more oxygen is produced, the faster photosynthesis had to run in the plant to produce this oxygen. (More info in Conclusion)

2. Describe, *in detail*, the part or parts of this investigation YOU personally designed.

I thought of what wavelengths of light we should investigate. I made sure that the choices were evenly spread over the color spectrum, to insure that most of the wavelengths of light are tested. I also designed the amount of time we should spend on each test counting bubbles, which was 3 minutes. I had to make sure that the time was long enough to get accurate test results while not so long that precious lab time is wasted.

3. Describe how the scientific concept you investigated in this component is related to a real-world issue or personal experience (you may include issues that affect society or the environment).

Every time we eat something, like hamburgers, hot dogs, spinach, and other foods, we are either directly or indirectly consuming the products of photosynthesis. It's obvious that when we eat plants, we are eating plants that nurtures itself with the glucose produced in photosynthesis. When we eat animals, those animals may have eaten other animals which may have eaten other animals which finally may have eaten the plant. The point is that if photosynthesis ceased to exist, there be no plants, which means there will be no animals because they feed on the non-existing plants and life on earth would just not be possible. Photosynthesis is a process we take for granted.

BEST COPY AVAILABLE

GSE Self-reflection Sheet: Problem-solving Investigation (cont'd)

4. Describe how working with others on this investigation helped to increase your understanding of science.

Before the lab, every single one of us was confused in their own way. When we got together and started discussing about it, our knowledge started to fill each other's gaps until we had the whole scientific concept of photosynthesis in our minds. Personally, I had no idea why the output of oxygen would tell us the rate of photosynthesis within the plant. I just knew the basic fact that the more oxygen produced, the faster photosynthesis is occurring. Now, we all know that the plant is not really trying to produce oxygen, but is producing glucose to nurture itself and oxygen just happened to be a waste product we need to survive produced while producing the glucose the plants need to survive.

5. What did you conclude from the investigation? Was the conclusion the same as or different from what you expected? Describe how your observations and data support your conclusions.

We concluded that different wavelengths of light do indeed effect the rate of photosynthesis. Our data shows oxygen being given off from the plant faster when red or white light was shined on the plant. We concluded that in the presence of red or white light, photosynthesis was occurring faster because the rate of oxygen being released indicated the speed of photosynthesis. We weren't surprised that white light was so effective because white light, unlike other colors, is not just one wavelength of light, but the whole spectrum of wavelengths of light combined. From the white light, the plant had access of every wavelength of light. We also found that in the presence of green light, the rate oxygen was being released was extremely slow compared to the other wavelengths of light, which means the rate of photosynthesis was also extremely slow. We conclude this is so because green is refracted from the leaves. When green light is shown on the plant, the plant probably reflects it back out instead of absorbs it because we could see the refracted green light. This is why plants are green. The results we saw was mostly what we expected because before starting the lab, we had the chance to get a lot of background information.

BEST COPY AVAILABLE

COMMENTARY

The student exemplar displays an understanding of photosynthesis and the detailed processes that affect its rate and products, although minor errors and omissions are present. The work provides an insightful description of the everyday relevance of photosynthesis in real-world terms. A clear student-designed problem, purpose, and investigation methodology are described in detail with appropriate data and results provided. Conclusions are appropriate and consistent with the data obtained by the student. The work shows the use of conceptual knowledge to guide the investigation and ultimately the results, confirming the students hypothesis — an essential aspect of scientific inquiry. The student makes recommendations for future steps to investigate the topic and concludes the work with an error analysis. While aspects of the inquiry can be improved, namely, studying aspects of photosynthesis in which the student cannot predict the results, the work shows strong evidence of a scientific understanding and the use of inquiry to solve problems.

notes

References

Bork, D., Martin, J., Bybee, R., Collins, A., Scott, E., and Yager, R. 1995. Introduction to the Secondary School Biology Domain Reference Project: A Framework for Biology. Chicago, IL: National Opinion Research Council.

California Department of Education. 1990. The California Science Framework. Sacramento, CA: California Department of Education.

California Department of Education. 1993. Golden State Examinations Science Portfolio Teachers Guide. Sacramento, CA: California Department of Education.

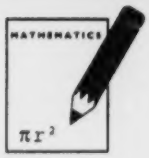
National Research Council. 1995. National Science Education Standards. Washington, DC: National Academy Press.

New Standards. 1995. Performance Standards Consultation Draft. Pittsburgh, PA: National Center on Education and the Economy.

Preliminary Draft: Illinois Academic Standards, Volume Two, State Goals 11-18, Science, June 1996.

State Collaborative for Assessment and Student Standards. 1995 working draft. SCASS Science Portfolio Handbook. Washington, DC: Council of Chief State School Officers.

State Collaborative for Assessment and Student Standards. 1996 working draft. SCASS Science Portfolio Implementation Guide. Washington, DC: Council of Chief State School Officers.



notes

Content and performance standards in mathematics describe goals for students at multiple grade levels. Many states have the same content and performance standards for students K-12. Benchmarks and achievement levels clarify the mathematical expectations for students at different grades. Assessments and their appropriate scoring guides may take many different formats. For example, some methods yield information about the student's knowledge and skills, while other methods focus on larger applications and nonroutine problem solving.

State Content Standard

Students use data collection and analysis, probability and statistics in problem-solving situations and communicate the reasoning used in solving these problems.

State Performance Standard

- ✔ Students will systematically collect, organize, and display data using charts, graphs, and tables; and
- ✔ analyze data to make valid inferences, arguments, and decisions.

State Assessment Task

Students at fourth grade level continue to read and create simple tables, charts, and graphs. They design surveys to gather data to answer specific questions. Fourth graders tabulate, organize, and display data in pictographs and bar graphs or tables and charts to make decisions and predictions based on their information.

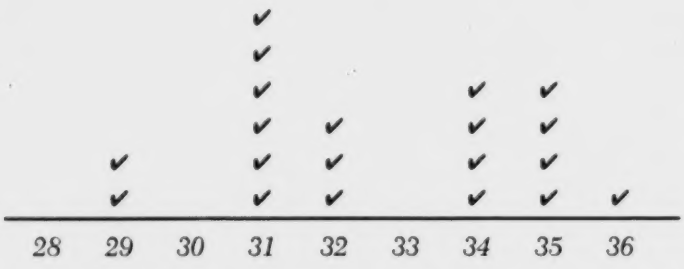
Fruit drinks may be purchased for 45¢ in the drink machine. The machine will not take pennies; it will take only quarters, dimes, and nickels. How many different ways can a student combine these coins to purchase a fruit drink using exact change?

A *5 ways* B *7 ways* C *8 ways* D *10 ways*

The correct answer is B. For a closer look at students' thinking, teachers use the same question and, rather than provide answer choices, have students make a table or chart to show all of the possible combinations of quarters, dimes, and nickels.

State Assessment Task

Marc and Jodi counted the number of crackers in their snack packs each day at lunch for three weeks. They recorded the number of crackers in each bag on this line plot.



Predict how many crackers you are likely to find in a snack pack just like the ones they have opened. Explain at least two reasons why you predict this number.

notes

Scoring Guide

4 Student gives number within the range of 29-34 and makes two or more supporting statements for that number. Student may go further and discuss why some other number may also be possible. Student discusses both range and measures of central tendency as part of the explanation.

3 Student gives number within the range of 28-37 and gives two supporting statements that relate to the data in the line plot. Discussion includes information about either range or measures of central tendency, but does not include both.

2 Student gives number within the range of 28-37 and makes one or more supporting statements, but does not clearly relate the number to measures of central tendency or the range.

1 Student gives number within the range of 28-37, but does not give reasons to support the answer, gives reasons that the data in the line plot do not support, or gives number outside the range.

Classroom Assessment Task

notes

You have been chosen by a local department store to make recommendations to the store's buyers about the products that the store should have to sell to its fourth-grade customers. You will make an oral presentation in 1 week to the store officials, using charts or graphs to support your recommendations.

A Choose a group of products such as shoes or toys and figure out what information you need in order to make recommendations to the store about the products you have chosen.

B Gather the information through an investigation that you design, organize the data, and make displays to use in your presentation.

You will be judged on how clear your ideas are, how well you planned and carried out your investigation, and how persuasive your presentation is.

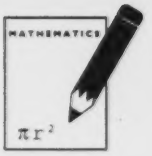
Scoring Guide

4 DESIGN, DATA COLLECTION, DISPLAYS

- ✔ Overall plan is evident; plan is complex and links each step in process to end results. Questions result in detailed information.
- ✔ Data have been collected systematically and accurately; adjustments have been made in questions so that information is full and complete.
- ✔ Sample population and sample size reflect work beyond minimums needed for task.
- ✔ Multiple, appropriate data displays give details; charts and graphs communicate information to audience.

4 INTERPRETATION, RECOMMENDATIONS, PRESENTATION

- ✔ Presentation is organized, and visuals are clear. Student describes the investigation and the processes used to conduct it.
- ✔ Delivery is engaging. Sentence structure is correct. Eye contact is made with audience; voice is clear and volume appropriate.
- ✔ Links are explicitly made to connect the steps taken to complete the task with the resulting data and the recommendations being made.



notes

- ✔ Data are interpreted logically, and recommendations are supported by data. Details are pointed out to audience as recommendations are made and alternatives are discussed.
- ✔ Questions from the audience are clearly answered with specific information.

3 DESIGN, DATA COLLECTION, DISPLAYS

- ✔ Student has plan for collecting data and uses appropriate sample population and sample size. Student revises questions as needed.
- ✔ Data displays (one or more) are appropriate for type of data and for desired outcome.
- ✔ Information is recorded and organized accurately; charts and graphs are labeled appropriately and accurately.
- ✔ Minor errors may exist in process or product, but they do not negate overall results.

3 INTERPRETATION, RECOMMENDATIONS, PRESENTATION

- ✔ Student describes investigation and talks about the process, but may not give specific details.
- ✔ Eye contact, clarity of speaking voice, and use of appropriate sentence structure are evident throughout most of the presentation.
- ✔ Statements refer to information in data display(s); appropriate recommendations are made, but are presented without details.
- ✔ Results of investigation are given, but may not include references to processes conducted. Interpretations are appropriate and accurate, but may not be fully explained.

2 DESIGN, DATA COLLECTION, DISPLAYS

- ✔ Data collection is not well organized; errors in recording, tabulating and/or displaying data lead to unsupported or confusing conclusions.
- ✔ Questions may relate to task, but are asked in a manner that makes it difficult to use the data. Sample population may not be representative of total population; sample size may be too small.

notes

- ✔ Data are presented in a single display, but display may lack clarity and/or accuracy. Labels may not be correct

2 INTERPRETATION, RECOMMENDATIONS, PRESENTATION

- ✔ Student states task, but fails to fully describe the investigation. Evidence of organization and preparation for presentation may be weak.
- ✔ Delivery is uneven. Eye contact, volume of voice, and use of appropriate sentence structure vary.
- ✔ Presentation may ignore data displays or may refer to them as incidental rather than as important to the conclusions.
- ✔ Recommendations may address the original task, but do not reflect the investigation conducted by the student.
- ✔ Inferences may be inappropriate and/or inaccurate for the question(s) asked and the data collected.

1 DESIGN, DATA COLLECTION, DISPLAYS

- ✔ Student demonstrates little understanding of how to accomplish task; no overall plan is in evidence. Task may be incomplete.
- ✔ Data collection is haphazard, inappropriate, and/or inaccurate.
- ✔ No appropriate data displays are created to communicate with audience; charts or graphs may be inaccurate.

1 INTERPRETATION, RECOMMENDATIONS, PRESENTATION

- ✔ Student fails to describe investigation or explain why it is important.
- ✔ Delivery is difficult to follow. There is no evidence of practice and/or other preparation for presentation.
- ✔ Recommendations do not address the task or are not supported by data. Recommendations may reflect personal opinion only.
- ✔ Student is unable to answer audience's questions.

Benchmarks and Achievement Levels

notes

The manner in which students complete tasks, the extent to which they accomplish stated goals, the consistency in their demonstrations of understandings and their skills are all indicators of their achievement of the standards. While scoring rubrics denote performance on specific assessments, determining achievement levels may involve gathering evidence from multiple sources over time. Different scales are used by different states.

Frequently, samples of students' work are used to illustrate the various levels of achievement of the same standard. These samples are benchmark papers to help bring consistency to the evaluation of all students' work.

Performance at different achievement levels has implications for the classroom. When goals are consistent for all students, the instructional needs of students who consistently perform at lower achievement levels may include additional time, modified learning opportunities, and special assistance. Likewise, students demonstrating high levels of achievement may move through developmental sequences more rapidly than their peers and attain content goals at earlier ages.

Descriptions of achievement levels provide indicators of performance at multiple levels. For example, the levels may be described as basic, proficient, and advanced. Achievement may be described along a developmental continuum from novice to expert performance. Individual tasks and instruments that serve as evidence for performance at different achievement levels may have scoring categories that are labeled in many different ways. For example, these labels may be advanced response, proficient response, nearly proficient response, minimal response, and nonscorable response.

Benchmarks and Achievement Levels

notes

LEVEL IV may be defined as performance that is beyond the age or grade level expectations and involves more difficult materials and greater independence; "expert" may be an appropriate label.

GENERAL LEVEL DESCRIPTORS

- ☑ Student works independently with confidence and initiative.
- ☑ Student applies problem-solving strategies in creative ways.
- ☑ Student analyzes and synthesizes, justifies and elaborates responses.
- ☑ Student makes critical judgments and understands advanced concepts.
- ☑ Student makes applications and extensions beyond grade level.
- ☑ Student consistently performs beyond grade level expectations.

COLLECTING, ORGANIZING, DISPLAYING, AND USING DATA

- ☑ Student designs and carries out complex data investigations, creating appropriate questions, determining sample population, and gathering data accurately; plan includes clear connections of steps with ultimate purpose.
- ☑ Student analyzes, synthesizes, and interprets data appropriately; student also justifies and elaborates interpretations.
- ☑ Student uses a variety of data displays, often presenting data in multiple ways to make information clearer; student relates data from different questions to final interpretations.
- ☑ Student includes range and measures of central tendency in discussions of results of investigations.

LEVEL III may be defined as consistent, accurate performance involving grade level materials and tasks; "proficient"may be an appropriate label.

GENERAL LEVEL DESCRIPTORS

- ☑ Student can apply, model, and explain concepts.
- ☑ Student applies strategies successfully in most situations.
- ☑ Student responds with appropriate answers or procedures.
- ☑ Student completes tasks accurately with minimal assistance.

notes

- ✔ Student is confident and takes appropriate risks.
- ✔ Student is consistent in application of skills and processes.
- ✔ Student exhibits proficiency with grade level tasks.

COLLECTING, ORGANIZING, DISPLAYING, AND USING DATA

- ✔ Student creates plan for data investigations, requesting assistance as needed.
- ✔ Student forms questions to gather information and revises them if necessary.
- ✔ Student collects, organizes, and displays data accurately and appropriately.
- ✔ Student identifies range and measures of central tendency.
- ✔ Student explains decisions in creating graphs and charts; student relates information in data displays to interpretations.

LEVEL II may be defined as uneven or inconsistent performance that does not yet exemplify accomplishment of the standard; "apprentice" may be an appropriate label.

GENERAL LEVEL DESCRIPTORS

- ✔ Student shows some evidence of conceptual understanding, but displays confusion or misunderstandings at times.
- ✔ Student completes routine procedures, but has difficulty with complex tasks or unfamiliar situations.
- ✔ Student is inconsistent in demonstration of skills.
- ✔ Student needs additional time and opportunities.
- ✔ Student work lacks enough detail to be considered proficient.

COLLECTING, ORGANIZING, DISPLAYING, AND USING DATA

- ✔ Student needs some assistance in planning and carrying out data investigations; student is successful when assisted.
- ✔ Student uses a limited number of data displays, may require reminders about appropriate labels, and may be inconsistent in accuracy.
- ✔ Student describes data using range and measures of central tendency if definitions are provided.
- ✔ Student's data investigations are not detailed and thorough.

LEVEL I may be defined as minimal performance demonstrating limited understandings; "novice" may be an appropriate label.

notes

GENERAL LEVEL DESCRIPTORS

- ☑ Student shows limited evidence of conceptual understandings.
- ☑ Student has very limited use of strategies.
- ☑ Student frequently responds to queries with inappropriate answer and/or procedure.
- ☑ Student rarely completes tasks appropriately and accurately.
- ☑ Student needs frequent assistance and guidance.

COLLECTING, ORGANIZING, DISPLAYING, AND USING DATA

- ☑ Student has difficulty making a plan and carrying it out without assistance at each step.
- ☑ Student displays data in bar graphs, but does not use a variety of data displays; graphs often lack appropriate labels.
- ☑ Student confuses measures of central tendency.
- ☑ Student's statements about graphs may not be supported by the data.

Instructional Activities

notes

MANTEO TO MURPHY

Instructional goal: Students will plan an experiment and gather data to answer a question. In the process, they will explore mean, median, and mode as ways of describing "typicalness."

Related learning experiences: Students will learn to work in teams of four, keep a log that describes steps in the investigation, reach consensus in order to complete an assignment, and make a persuasive presentation to the class.

QUESTION TO BE INVESTIGATED:

How long would it take the typical fourth-grade student to walk the length of North Carolina from Manteo to Murphy?

PLANS — DAY 1: Have each group find out (a) how long a mile is and (b) the distance from Manteo to Murphy. Groups then make a specific plan for determining the length of time it would take each student in the class to walk a mile. The plan should indicate where, when, and how the group will collect the data. Groups are to make the recording sheets that they will use while gathering the data. Remind students that the distance around the track on the playground is one-fourth of a mile and there are 28 students in the class. After groups present their plans to the class, discuss the need for accuracy in recording data. Have students practice with stopwatches.

PLANS — DAY 2: Students collect data during outside time. In class, review and compare measures of central tendency that students used in earlier investigations. Have the class as a whole make lists of advantages and disadvantages of using each measure in this specific situation. Give groups time to organize their data and figure out an answer to the question. Be sure they are prepared to talk about how the group will define "typical."

PLANS — DAY 3: Groups plan a presentation with charts or graphs to persuade others in the class that their conclusions are accurate and appropriate. While groups are making charts and getting ready for their presentations, talk with individual students to see who needs more work on measures.

PLANS — DAY 4: After groups make their presentations, have the class vote on what they think is the best answer to the question. Talk about how the charts and graphs helped to give the audience information and how the ways in which the groups made their calculations varied.

Instructional Activities

notes

FOLK TALE HEROES AND HEROINES

Instructional goal: Students will recognize that circle graphs (pie charts) show the relationship of the parts to the whole and to each other. They will explore one way to create a pie chart and will practice interpreting charts created by others.

Related learning experiences: Students will learn to create a log of activities, use estimation and rounding, and forge links to literature by imagining activities of characters from folk tales and favorite classics.

QUESTIONS TO BE INVESTIGATED:

How do you spend your day? How would various characters from folk tales and stories spend their days? How can we show this information on a circle graph?

PLANS — DAY 1: Students use a 24-hour chart that has half-hour segments to log how they spend their time. As students are recording their personal data, discuss the differences in weekends and school days and ways to estimate and compensate for activities that take less than half an hour. The next step is to organize the data (i.e. clump all of the meal and snack times together to form a category that might be labeled "eating"). Working with a partner, students help each other organize their individual data and begin to compare their typical days.

PLANS — DAY 2: Talk about how students can represent their data on a circle graph. Model the process using personal data as the class carries out each step. Give each student a large paper circle. Talk about how they may divide the circle to help them determine 1-hour segments. If students fold their circles in half three times the result will be eight segments, each representing 3 hours. Show (teacher models) students how they can subdivide one of these 3-hour sections into 1-hour segments if necessary. Have students record their own data by coloring the sections, mounting their pie chart on another sheet, and creating a key to explain the colors. Have students title their charts and display them so that they can be compared.

PLANS — DAY 3: Have each student, with a partner, choose a fictional character or folk hero and create a circle graph to show how the character might have spent a typical day. For example, Blackbeard might have slept for 7 hours, eaten for 2 hours, gone fishing for 5 hours, worked on burying his treasure for 4 hours, practiced with his sword and pistol for 2 hours, and sailed around for 4 hours. Use the graphs to create a "Mystery Person" bulletin board where students can try to guess the identity of the characters from the activities shown on the circle graphs. Assess individual students' understanding of circle graphs by having them interpret different graphs from the display.

48

Instructional Activities

notes

TEAMS WE LIKE TO SUPPORT

Instructional goal: Students will create a questionnaire; survey classmates, other students, and adults in the community; and enter the information into a database. Using the class computer(s), they will learn to sort data in a variety of ways and compare the results of the sorts.

Related learning experiences: Students will learn to plan based on previous experiences, collect and record data accurately, use keyboarding skills, write letters, and respect differences of opinion.

QUESTIONS TO BE INVESTIGATED:

What are the favorite teams of people in our area? What are the differences, if any, for different sports by age group or by gender?

PLANS — DAY 1: Students brainstorm the questions they need to ask on their survey to gather data in order to answer the questions that have been posed. Help students link previous experiences with sampling to the decisions that the class makes for this investigation. After the class discussion, have students work in groups to carry out different tasks. One group creates the questionnaire; another group works with the teacher to set up categories for the database; a third group creates a timeline for the investigation and writes guidelines for administering the survey; and a fourth group writes letters to the teachers in other classes that are being surveyed to make arrangements to collect data from their students. At the end of the class, the students come together to be sure that the work from each group fits with the work of other groups.

PLANS — DAY 2: Students gather data in their own classroom, in other classes, and at home that evening. As the data are collected, students take turns entering the data into the computer. (Because students do different things during the math period, day 2 allows for follow-up to previous work, informal assessment and time for students to work on their portfolios.)

PLANS — DAY 3: After remaining data have been entered, students can sort the information in various ways. Give each pair of students copies of the sorted data and have them discuss the results of the surveys. Pairs of students may join other pairs to talk about their observations. Have students respond individually to this prompt: "Based on our class investigation, I think the favorite team of people in this area is _____. I think this because _____."

EXTENSION: Using the computer, students can explore some "what if" questions related to sampling: What if we entered the data incorrectly for 10 students? How might that change our results? What if we surveyed only 10 adults, but questioned 75 students? What if we asked all girls and no boys? What if we asked about specific sports?

State Content Standard

Students understand and use various types of operations and computational techniques, linking concepts and procedures to solve problems.

State Performance Standard

- ✔ Students will demonstrate proficiency with basic facts;
- ✔ model, explain and use the four basic operations of addition, subtraction, multiplication, and division to solve problems; and
- ✔ select and apply appropriate algorithms and computational techniques such as estimation, mental arithmetic, paper and pencil, calculators, and computers.

Computation is an important part of mathematics and of our daily lives. It best serves us when computational skills are linked with "number sense" and "operation sense." It is a tool used throughout mathematics: we measure and compute, gather data and tabulate results, state percentages of and calculate selling prices, and look at the clock and estimate when to leave home in order to arrive on time.

notes

Classroom Assessment Task

notes

Lynn turned in this paper. Circle the problems that are incorrect. Write a note to Lynn explaining why some of the problems are wrong and how to correctly solve them.

A

$$\begin{array}{r} {\scriptstyle 1} \\ 7062 \\ +\ 1594 \\ \hline 8656 \end{array}$$

B

$$\begin{array}{r} 816 \\ -\ 425 \\ \hline 411 \end{array}$$

C

$$\begin{array}{r} {\scriptstyle 1} \\ 709 \\ +\ 185 \\ \hline 984 \end{array}$$

D

$$\begin{array}{r} {\scriptstyle 81012} \\ 6\not{9}\not{0}\not{2} \\ -\ 4594 \\ \hline 2318 \end{array}$$

E

$$\begin{array}{r} {\scriptstyle 1} \\ 3080 \\ +\ 2547 \\ \hline 5627 \end{array}$$

F

$$\begin{array}{r} {\scriptstyle 1\ 1} \\ 264 \\ +\ 4579 \\ \hline 4842 \end{array}$$

Scoring Guide

ADVANCED RESPONSE Student recognizes correct and incorrect responses, marking them appropriately, writes a complete explanation of why the problems are incorrect and how they should be corrected, identifies errors by categories such as number facts or regrouping mistakes.

DEMONSTRATES UNDERSTANDING Student recognizes correct and incorrect responses, marking them appropriately, and illustrates how problems should be corrected. Explanation of the mistakes may not be precise, but each error is addressed.

INCOMPLETE RESPONSE Student recognizes some of the incorrect responses. Student shows how errors should be corrected, but gives only limited explanation of error. Student may identify all of the incorrect problems, but does not explain why any are incorrect.

MINIMAL OR INADEQUATE RESPONSE Student does not recognize incorrect responses. Student may mark correct answers as incorrect, giving an inaccurate explanation.

notes

Each Book Had Three Pages: A Grade Three English Language Arts Classroom Performance Model

Developed by Maggie Alexander, Maple Lane Elementary School; Lisa Best, Lulu Ross Elementary School; Susan Gilmore, Smyrna Elementary School; Marianne Kellner, Linden Hill Elementary School, for Delaware Department of Public Instruction

State Content Standards

- ☑ Construct, examine, and extend the meaning of literary, informative, and technical texts through listening, reading, and viewing.
- ☑ Access, organize, and evaluate information gained through listening, reading, and viewing.

State Performance Standard

Requisite Prior Instruction (model instructional unit available): In order for the third-grade student to successfully complete this task, the educational program (K-3) must have offered sufficient and appropriate instruction and practice in the following bodies of knowledge, abilities, and skills:

- ☑ Reading nonfiction, informative text
- ☑ Reference and research skills
- ☑ Note taking

NOTE: It must be emphasized that the requisite instruction and practice must not occur only in the weeks immediately preceding the task, or even only in the school year in which the task is to be administered.

Performance Task

The following task is an example of an instructional task with an embedded assessment. While the task itself integrates English language arts and math, only English language arts will be assessed. The task requires students to "construct, examine, and extend informative text" and use the information gained to "make decisions, solve problems, complete

notes

tasks, and create products." Students will also "access, organize, and evaluate information . . . with direct teacher assistance," as specified in the content standards, demonstrating an overall understanding of the informative text they choose as they select and organize information in order to produce a short book patterned after *Each Orange Had 8 Slices* by Paul Giganti, Jr.

Scores are based on the relevance and accuracy of information gained from students' research, and on the students' ability to organize information and present it in a prescribed format.

Final task activities and products require students to select a topic of interest to them and find information on that topic. With direct teacher guidance, students organize and present information.

NECESSARY MATERIALS

- ☑ One copy of *Each Orange Had 8 Slices* by Paul Giganti, Jr. (Greenwillow). (Multiple copies or a big book would be helpful, but not necessary.)
- ☑ One copy of *A Swim Through the Sea* by Kristin Joy Pratt (Dawn Publications).
- ☑ Informational books on a wide variety of topics.
- ☑ Paper, pencils, markers, crayons, construction paper, etc.
- ☑ Overhead projector, transparencies, and markers.
- ☑ One copy of a related book: *How Many Snails?* by Paul Giganti, Jr. (Greenwillow).

ADVANCE PREPARATION

Before beginning this task, the teacher should do the following

- ☑ Read the entire task.
- ☑ Gather necessary materials as listed above.
- ☑ Prepare the graphic organizer for use in Task Activity 2 (page 47).
- ☑ Prepare transparencies for Task Activity 3; make sufficient copies of all student activity forms.

Task Activities

notes

(Class time required: 11 45-minute sessions)

TASK ACTIVITY 1

Read *Each Orange Had 8 Slices* aloud to the students for their enjoyment, showing the pictures as you read. Read the book to the students a second time, but this time do not show the pictures. As each page is read, allow the students time to solve the problems and whisper their answers to the person next to them. Then show the picture so the students can check their answers. After completing the book, reread several pages and let the students tell how they solved the problem. Discuss the pattern of the book, which is a mathematical word problem followed by three questions. Discuss the format of the word problems in the book.

TASK ACTIVITY 2

Tell students that they will be researching a topic and using the information to produce a book patterned after *Each Orange Had 8 Slices*. Before being asked to produce an individual book, they will produce a class book about another topic. Sea life is being used as an example here.

Draw the graphic organizer shown in Student Page A (page 47) on the chalkboard or overhead projector. (If students have not used graphic organizers, explain that a graphic organizer is a picture representing information that is organized in a way that makes the information easier to understand. A web is one type of graphic organizer.)

Draw the circle and explain that this is the area where the topic (category) is listed. Draw several spokes from the circle outward to illustrate where subtopics (or subcategories) are listed. Have students brainstorm a list of subtopics and record them on the spokes (e.g., crustaceans, mammals, shellfish, fish, birds, plant life). Divide the class into as many groups as there are subtopics. Have each group choose a subtopic, read about it and record examples of the subtopic on the spokes.

The use of the graphic organizer helps students visualize their thoughts while encouraging them to put their initial ideas on paper so as not to forget them. This type of thinking/planning is essential as the ultimate quality of the students' books will hinge on careful identification and organization of appropriate information.

notes

TASK ACTIVITY 3 (2 SESSIONS)

Read several pages from *A Swim Through the Sea*. Select a page from the book to show on the overhead. Model how to pick out the important information and take notes on it. Show another page from the book on the overhead. Have students take turns coming to the overhead, choosing important information and underlining it with markers. Have the students discuss whether the underlined information is important and why.

This modeling and guided student practice allow the students to become familiar with the idea of identifying pertinent information. The discussion of whether the selected information is important allows students to develop an understanding of the underlying thought processes. Otherwise, this might become simply an exercise in underlining words on a transparency. Doing this as a class activity allows each student to participate and offers each student immediate feedback.

Select a third page from *A Swim Through the Sea* and show it on the overhead. Give each student a copy of Student Page B — Note Taking Sheet (page 49). Have students complete one section of this form using the information on the overhead. Have the students share and compare responses.

TASK ACTIVITY 4 (3 SESSIONS)

Reread *Each Orange Had 8 Slices*. Remind the students that they will be making a class book patterned after this book. Using the information recorded from *A Swim Through the Sea* (on Student Page B — Note Taking Sheet), model how to use this information to complete Student Page C — Planning Sheet (page 51) to write a page for the class book.

After you have completed a page, divide the class into the small groups formed in Task Activity 2. Instruct the students to choose one of the examples listed under their group's subtopic, find some information relating to it and take notes using Student Page B — Note Taking Sheet. Tell them that they will use this information to complete Student Page C — Planning Sheet, which will contain the rough draft of their page. Remind students that they should proofread and edit their work before writing the final draft and illustrating their page so that the illustration matches the text. When all pages are complete, students should choose a title and select students to make the cover and title page. The book should be shared when completed.

notes

While cooperative practice is an important component of this task and helps prepare the students to work individually on the book that will be assessed, these sessions should not be the only times during grade 3 — and in grades K-2 — when students have instruction and practice in reading informative text and identifying important or key information. Activities like this can be part of social studies, science, or mathematics instruction.

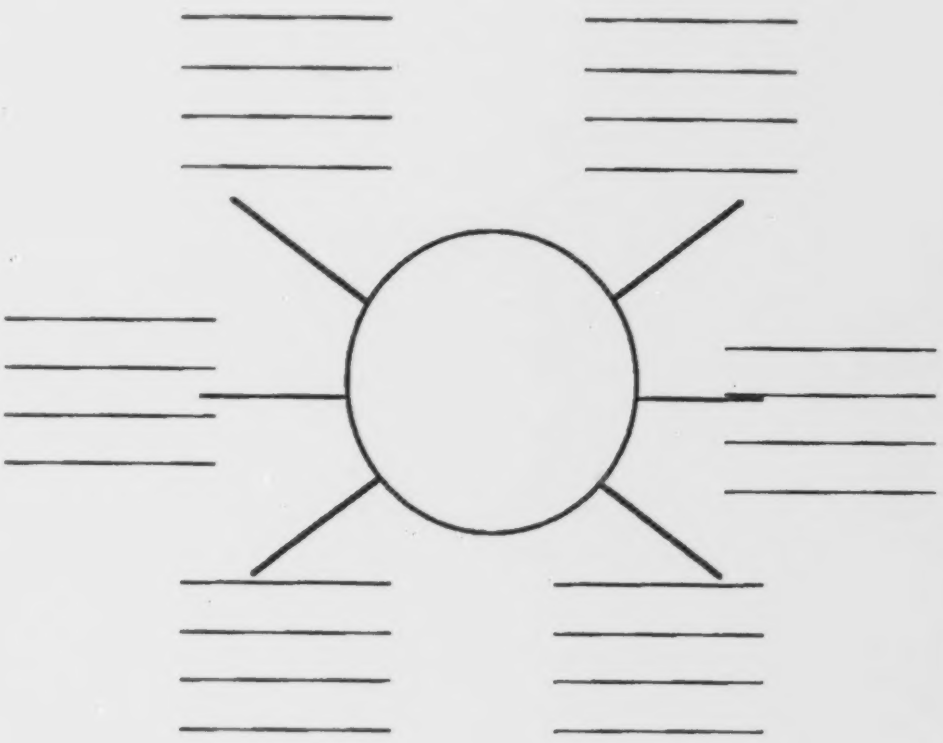
TASK ACTIVITY 5 (3 SESSIONS)

Reread the class book. Remind students that they did research on a topic and then used the book *Each Orange Had 8 Slices* as a model to write the class book. Tell the students that they will be using the same process to write an individual book on a topic of their choice. Have the students brainstorm a list of possible topics. When the students have chosen their topics, distribute copies of Student Pages A, B, and C as needed and a copy of Student Page D — Steps to Create Your Book (page 53).

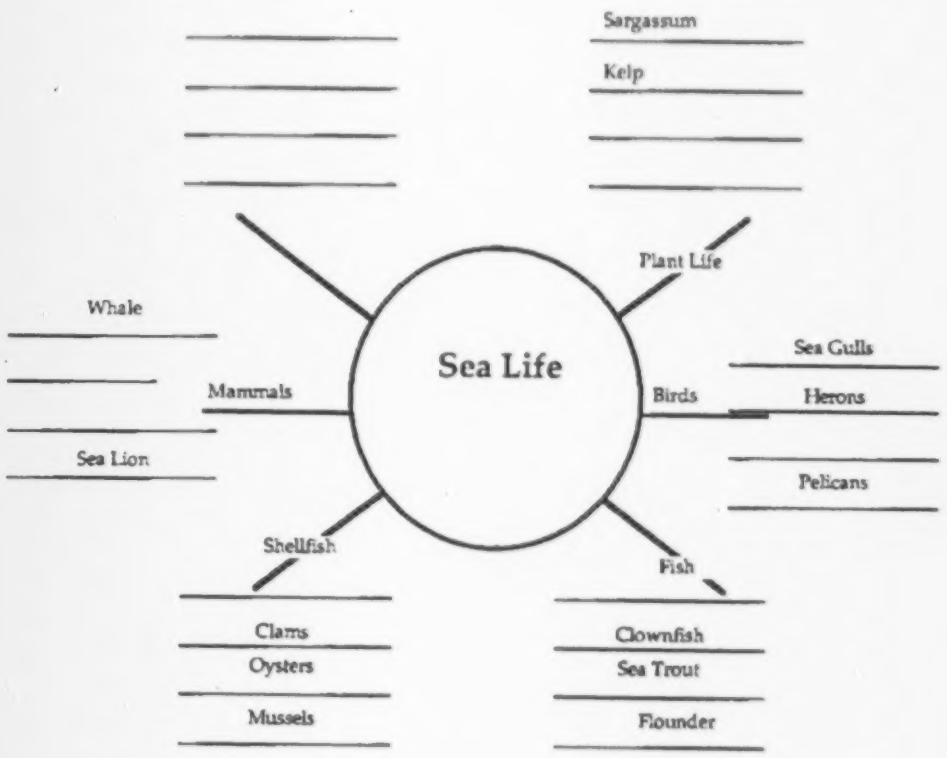
TASK ACTIVITY 6

After completing their books, students should read them and solve the problems that they created. As a culminating activity, each student chooses one page from his or her own book and shares it orally with their classmates, who will try to solve the problem.

TASK ACTIVITY 2: STUDENT PAGE A — GRAPHIC ORGANIZER



TASK ACTIVITY 2: EXAMPLE OF PARTIALLY COMPLETED GRAPHIC ORGANIZER



BEST COPY AVAILABLE

TASK ACTIVITY 3: STUDENT PAGE B — NOTE TAKING SHEET

Name: _____ Date: _____

Topic: _____ Subtopic: _____ Example: _____

Fact: _____

_____

_____

Additional Information: _____

_____

_____

Title: _____

Author: _____ Page: _____

Name: _____ Date: _____

Topic: _____ Subtopic: _____ Example: _____

Fact: _____

_____

_____

Additional Information: _____

_____

_____

Title: _____

Author: _____ Page: _____

TASK ACTIVITY 3: EXAMPLE OF PARTIALLY COMPLETED NOTE TAKING SHEET

Name: _____ Date: _____

Topic: *Sea Life* Subtopic: *Fish* Example: *Clownfish*

Fact: *The clownfish has a special relationship with the sea anemone. It lays its eggs in the anemone for protection. It also helps feed the anemone.*

Additional Information: *Clownfish has mucus on its scales so anemone can't sting it. White stripes for camouflage -- make it look like part of the anemone.*

Title: *A Swim Through the Sea*

Author: *Kristin Pratt* Page: *4*

TASK ACTIVITY 4: STUDENT PAGE C — PLANNING SHEET

A Number of and Name of Item	B Number and Characteristic or Association for Item in A	C Number and Characteristic or Association for Item in B
_____	_____	_____

Setting: _____

Narrative : _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Questions:		Answers:
1. How many _____	were there?	_____
2. How many _____	were there?	_____
3. How many _____	were there?	_____

TASK ACTIVITY 4: EXAMPLE OF PARTIALLY COMPLETED PLANNING SHEET

A Number of and Name of Item	B Number and Characteristic or Association for Item in A	C Number and Characteristic or Association for Item in B
3 Sea Anemones	*2 clownfish*	*6 white stripes*

Setting: *Aquarium*

Narrative: *Story problem: When I went to the aquarium with my class I saw 3 sea anemones. Each sea anemone had 2 clownfish hiding in it. Each clown fish had 6 white stripes.*

Questions:			Answers:
1. How many	*Sea anemone*	were there?	*3*
2. How many	*Clownfish*	were there?	*6*
3. How many	*White Stripes*	were there?	*36*

notes

TASK ACTIVITY 5: STUDENT PAGE D — STEPS TO CREATE YOUR BOOK

Directions: Check off each step as it is completed.

- ☐ Choose a topic.
- ☐ Locate *at least* two books related to that topic.
- ☐ After doing some reading, fill in the topic, subtopics, and examples on your graphic organizer (Student Page A).
- ☐ Takes notes on the note taking sheets (Student Page B).
- ☐ Select the information to be used in your book, and record it on the planning sheets (Student Page C).
- ☐ Complete a rough draft of three pages, sharing one fact per page, using the pattern of the book.
- ☐ Solve the problem as you complete each page, and record your answer (Student Page C).
- ☐ Proofread your work, and edit your text.
- ☐ Write your final draft.
- ☐ Illustrate each page to match the text.
- ☐ Make a title page and cover when all the pages are complete.
- ☐ Read your book and select the page to be shared.

notes

STUDENT PAGE E — STUDENT WORK SCORING CRITERIA

(This is what your teacher will look for in scoring your work.)

GRAPHIC ORGANIZER

- ☑ Subtopic is related to topic.
- ☑ Examples are related to subtopic.

NOTE TAKING/PLANNING SHEETS

- ☑ More than one sheet is completed.
- ☑ Selected points are important.
- ☑ Details are relevant.
- ☑ Information is accurate.
- ☑ Notes make sense.
- ☑ More than one source is used.
- ☑ Student has gathered enough information to complete a book.

BOOK

- ☑ Book is completed.
- ☑ Pattern is followed.
- ☑ Illustrations match the text.

Scoring Guide

(Should be made available to students as they work)

4 POINTS

GRAPHIC ORGANIZER

- ☑ Subtopic is related to topic.
- ☑ Examples are related to subtopic.

NOTE TAKING/PLANNING SHEETS

- ☑ More than one sheet is completed.
- ☑ Selected points are important.
- ☑ Details are relevant.
- ☑ Information is accurate.
- ☑ Notes make sense.
- ☑ More than one source is used.
- ☑ Student has gathered enough information to complete a book.

BOOK

- ☑ Book is completed.
- ☑ Pattern is followed.
- ☑ Illustrations match the text.

3 POINTS

GRAPHIC ORGANIZER

- ☑ Subtopic is related to topic.
- ☑ Examples are related to subtopic.

NOTE TAKING/PLANNING SHEETS

- ☑ More than one sheet is completed.
- ☑ Some of the selected points are important.
- ☑ Some details are relevant.
- ☑ Information is accurate.

notes

READING STANDARDS

notes

- ☑ Notes make sense.
- ☑ More than one source is used.
- ☑ Student has gathered enough information to complete a book.

BOOK

- ☑ Book is completed.
- ☑ Pattern is followed.
- ☑ Illustrations match the text.

2 POINTS

GRAPHIC ORGANIZER

- ☑ Subtopic and examples may be confused.
- ☑ Examples may be unrelated to subtopic.

NOTE TAKING/PLANNING SHEETS

- ☑ At least one sheet is completed.
- ☑ Several selected points are unimportant.
- ☑ Some details may be irrelevant.
- ☑ Information is accurate.
- ☑ Student may have used only one source.
- ☑ Student has gathered enough information to complete a book.

BOOK

- ☑ Book is completed.
- ☑ Pattern is not consistently followed.
- ☑ Illustrations do not always match the text.

1 POINT

GRAPHIC ORGANIZER

- ☑ Subtopic and examples may be confused.
- ☑ Examples may be unrelated to subtopic.
- ☑ Graphic organizer may be incomplete.

notes

NOTE TAKING/PLANNING SHEETS

- ☑ Sheet may be only partially completed.
- ☑ Many selected points are unimportant.
- ☑ Some details may be irrelevant.
- ☑ Some information is inaccurate.
- ☑ Student may have used only one source.
- ☑ Student may not have gathered enough information to complete a book.

BOOK

- ☑ Book is incomplete.
- ☑ Pattern is not consistently followed.
- ☑ Illustrations do not always match the text.

This scoring guide does not address such conventions of standard written English as spelling, grammar, and punctuation because this task does not assess student writing. Since the students are writing to a pattern, it is not appropriate to use this as an example of a writing assessment.

SAMPLE STUDENT WORK 1

Name: _____

STUDENT PAGE C - PLANNING SHEET

Topic: Horses Subtopic: Bronco

Fact: A bronco is a horse from the state of Mexico. The horse came with the Spanish explorers from Mexico.

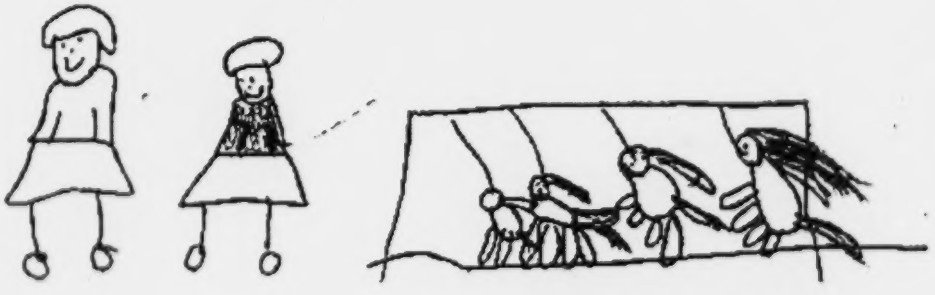
Related Information: The bronco is just like any other horse. Now they are not any broncos in the United States. They eat corn, hay, and drink milk.

A Number/Item	B Number/Characteristic or Association for A	C Number/Characteristic or Association for B
2 explorers	1 horses	1 bag of corn

Setting: Mexico

Narrative: On my way to Mexico, I saw 2 explorers. Each explorer had 2 hands. Each bronco had 1 bag of corn,

Questions: How many explorers are there? How many horses are there?
How many bags of corn are there in all?



BEST COPY AVAILABLE

SAMPLE STUDENT WORK I

SAMPLE STUDENT WORK I

Horses Horse food

On my way to the meadow, I saw 6 wild horses. Each wild horse was drinking out of 1 pail of water. Each pail of water had 1 lump of suga How many wild horses are there How many pails are there? How many lumps of sugar are th in all?

BEST COPY AVAILABLE

notes

LANGUAGE ARTS

notes

SAMPLE STUDENT WORK 1

Horses Gaits

On my way to the race, I saw 2 riders. Each rider had 1 horse. Each horse was practicing 10 gallops. How many riders are there? How many horses are there? How many gallops are there in all?

SAMPLE STUDENT WORK 1

Horses Branco

On my way to Mexico, I saw 2 explorers. Each explorer had 1 horse. Each horse had 1 bag of corn. How many explorers are there? How many horses are there? How many bags of corn are there in all?

BEST COPY AVAILABLE

70

LANGUAGE ARTS

notes

SAMPLE STUDENT WORK 1

Horses Breeding

On my way to the farm. I saw 2 horses. Each horse had 1 baby. Each baby had a piles of hay. How many horses are there? How many babies are there? How many piles of hay are there in all?

SAMPLE STUDENT WORK 1

Horses Colts

On my way to the countryside, I saw 2 colts. Each colt had 1 bottle. Each Bottle had 2 liters of milk in it. How many colts are there? How many bottles are there? How many liters of milk are there in all?

(2) (2)

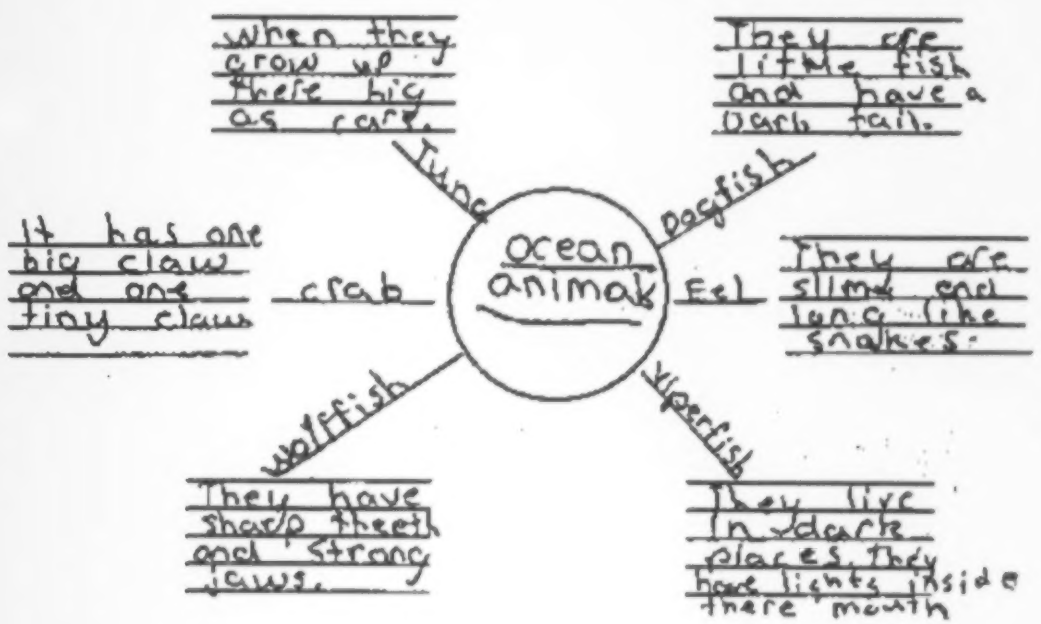
SCORE POINT 4

The graphic organizer is complete; topics and subtopics are appropriately related. Although only one Student Planning Sheet is reproduced here, more than one sheet was completed. These sheets contain some important information and relevant details that are reflected in the book (e.g., introduction of broncos by Spanish explorers). The book itself is complete and consistently follows the pattern; the illustrations match the text.

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 2

STUDENT PAGE A - GRAPHIC ORGANIZER


when they grow up there big as cars.
They are litle fish and have a Dark tail.
Tuna
Dogfish
ocean animak
It has one big claw and one tiny claws
crab
Eel
They are slimy and long like snakes.
wolffish
viperfish
They have sharp theeth and strong jaws.
They live in dark places, they have lights inside there mouth

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 2

Name: _____

STUDENT PAGE C - PLANNING SHEET

Topic: ocean animals Subtopic: yellow tail flounder

Fact: These fish are flat and have flat bodys

Related information: They are really yellowed tailed flounder beacuse there tail is yellow. They are ugle. They got to ugle eyes.

A Number/Item	B Number/Characteristic or Association for A	C Number/Characteristic or Association for B
2 yellow flouder	2 ugle eyes	3 dots

Setting: _____

Narrative: On my way to get my mom I saw a yellow flounder On the flounder there was two ugly eyes. On each ugly eye there were 3 dots.

Questions: How many yellow flounder there? How many dots are there?

How many ugle eyes are there in all?

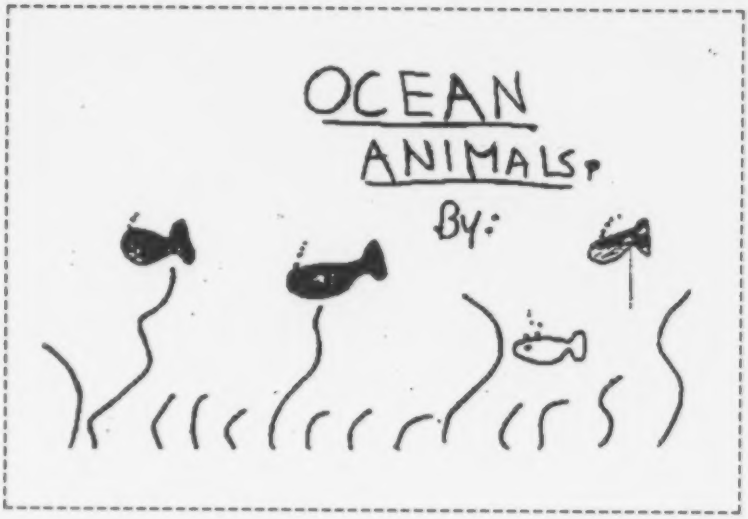
BEST COPY AVAILABLE

LANGUAGE ARTS

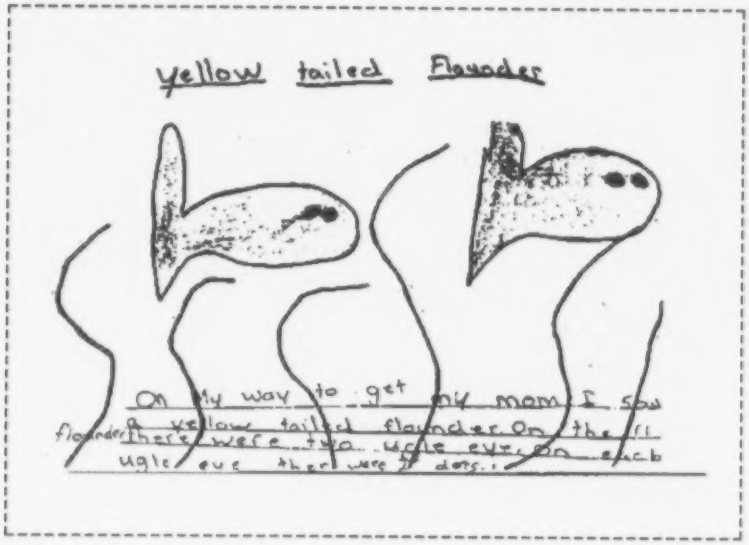
message

notes

SAMPLE STUDENT WORK 2



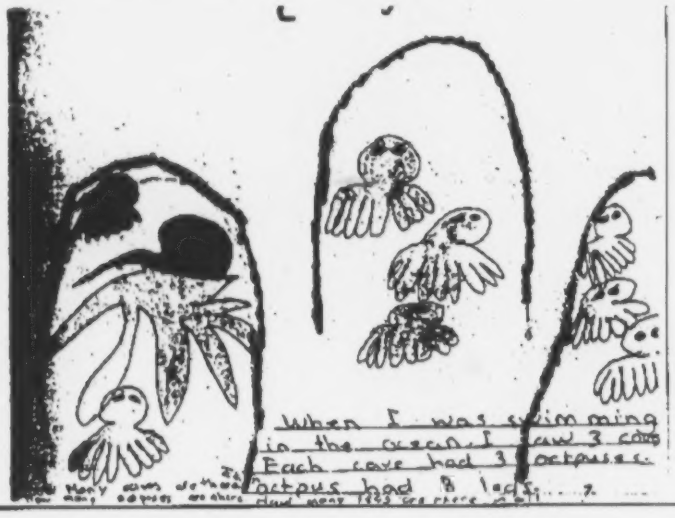
SAMPLE STUDENT WORK 2


Yellow tailed Flounder
On my way to get my mom I saw
a yellow tailed flounder. On the fl
flounder There were two ugly eyes on each
ugle eye there were 10 dots.

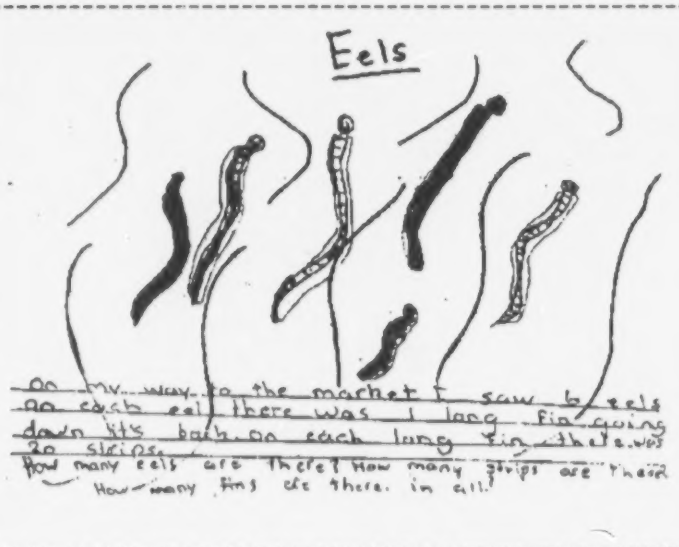
BEST COPY AVAILABLE

notes

SAMPLE STUDENT WORK 2

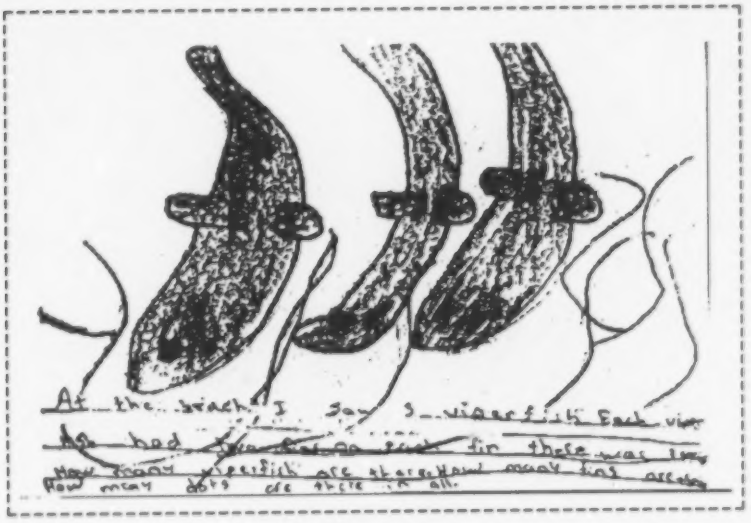


SAMPLE STUDENT WORK 2

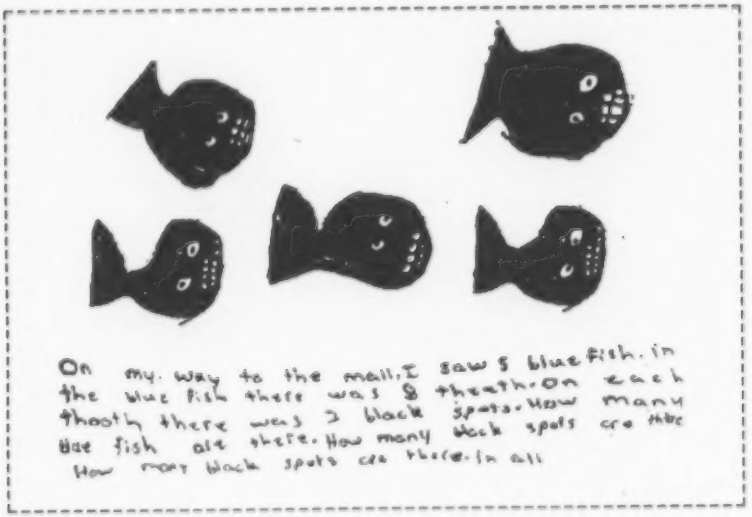

Eels
On my way to the market I saw 6 eels
On each eel there was 1 long fin going
down its back. On each long fin there was
20 stripes.
How many eels are there? How many stripes are there?
How many fins are there in all?

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 2


At the beach I saw 3 viper fish Each viper
How many viperfish are there. How many fins are there
How many dots are there in all.

SAMPLE STUDENT WORK 2


On my way to the mall, I saw 5 blue fish. In
the blue fish there was 8 teeth. On each
tooth there was 3 black spots. How many
blue fish are there. How many black spots are there
How many black spots are there in all

SCORE POINT 3

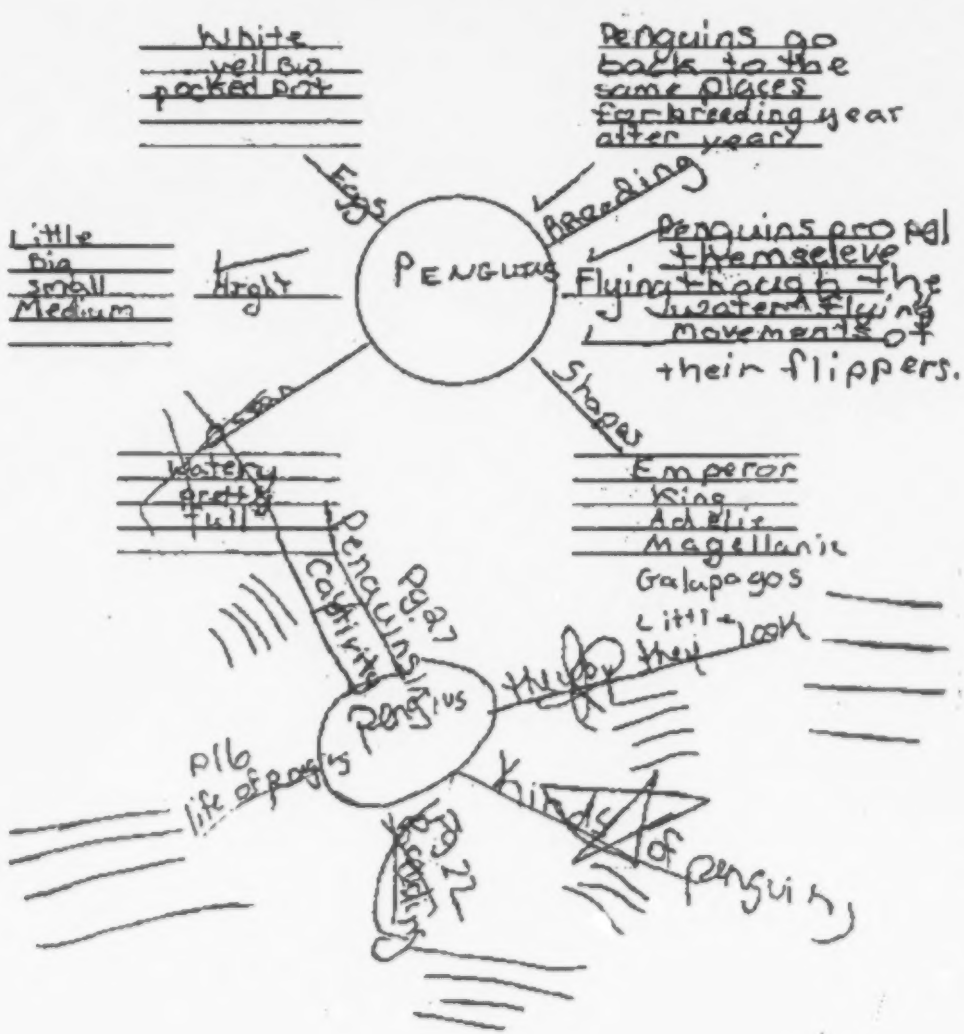
The graphic organizer is complete, topics and subtopics are appropriately related. Although only one Student Planning Sheet is reproduced here, more than one sheet was completed. These sheets contained some important information and relevant details (e.g., called yellow-tailed flounder because they are yellow) and some less relevant or important points (e.g., ugly eyes). The book itself is complete and follows the pattern; the illustrations match the text.

notes

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 3

STUDENT PAGE A - GRAPHIC ORGANIZER



BEST COPY AVAILABLE

SAMPLE STUDENT WORK 3

Name: _____

STUDENT PAGE C - PLANNING SHEET

1) Topic: Penguins Subtopic: Breeding

Fact: Penguins always go back to the same place to breed year after year. A colony of king penguins is called a rookery.

Related Information: Island. These penguins breed on islands and around the Antarctic continent. Penguins build their nests on the ground using rocks and stones or grass. Some nests in caves. Most penguins lay three eggs but rarely more than two chicks survive. Only one of every four.

A Number/Item	B Number/Characteristic or Association for A	C Number/Characteristic or Association for B
3 penguins	10 feathers	1 spot

Setting: (Look back on!)

Narrative: One day when I was walking around I saw a group of penguins. Each penguin had 10 feathers. Each feather had 1 spot.

Look on back for Story

Questions: How many penguins are there? How many feather are there?

How many spots are there in all?

A B C

3 penguins 10 pieces of hair 1 point at the end Setting: Igloos

Narrative:

I was on my way to the igloos and I saw 3 penguins. Each penguin had 10 pieces of hair. Each piece of hair had a point at the end.

Questions: How many penguins are there

How many pieces of hair are there?

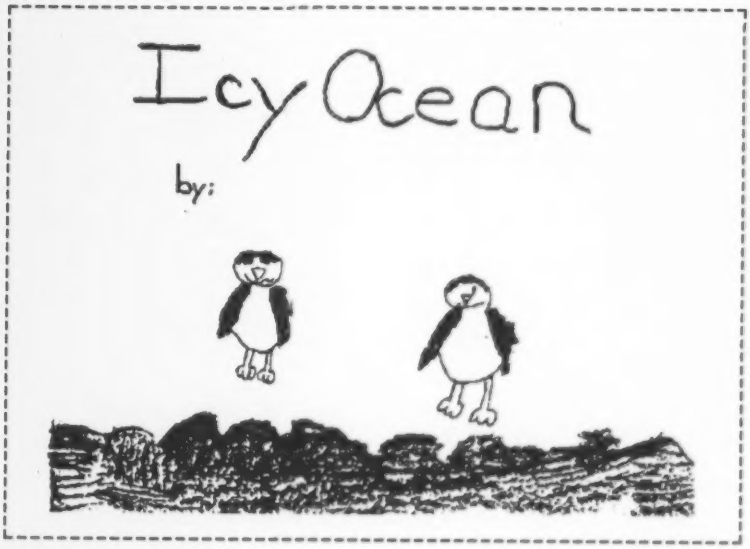
How many points are there in all?

BEST COPY AVAILABLE

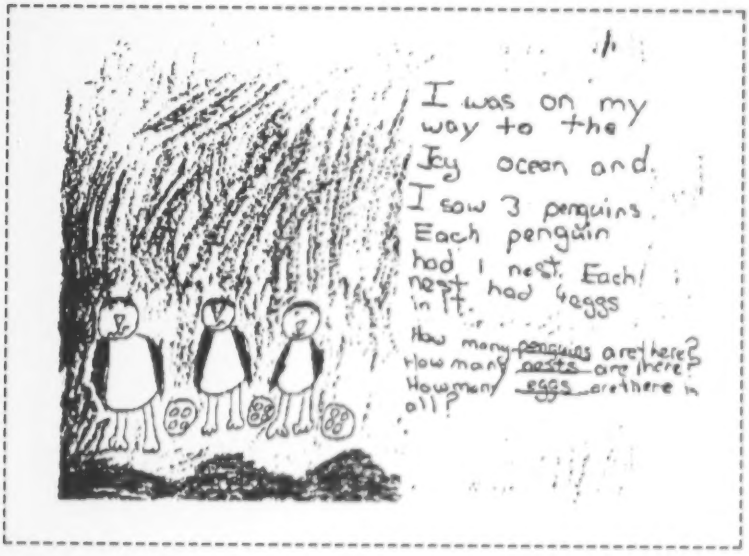
LANGUAGE ARTS message

notes

SAMPLE STUDENT WORK 3



SAMPLE STUDENT WORK 3


I was on my way to the Icy ocean and I saw 3 penguins. Each penguin had 1 nest. Each nest had 4 eggs in it.
How many penguins are there?
How many nests are there?
How many eggs are there in all?

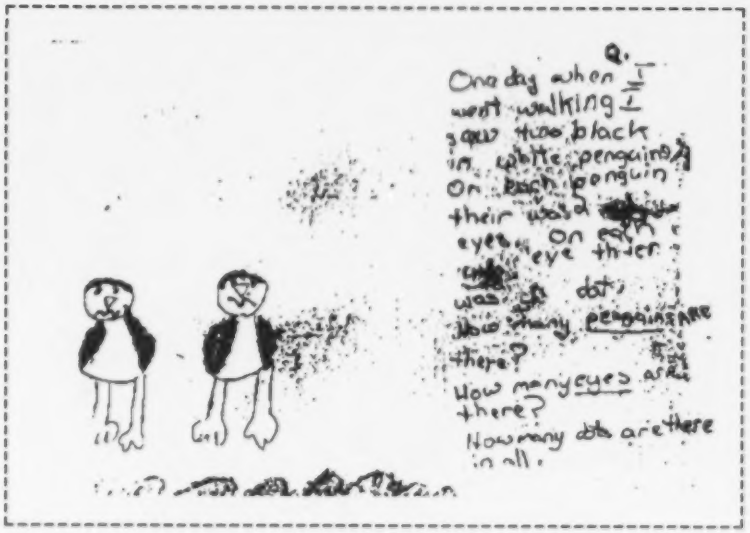
BEST COPY AVAILABLE

79

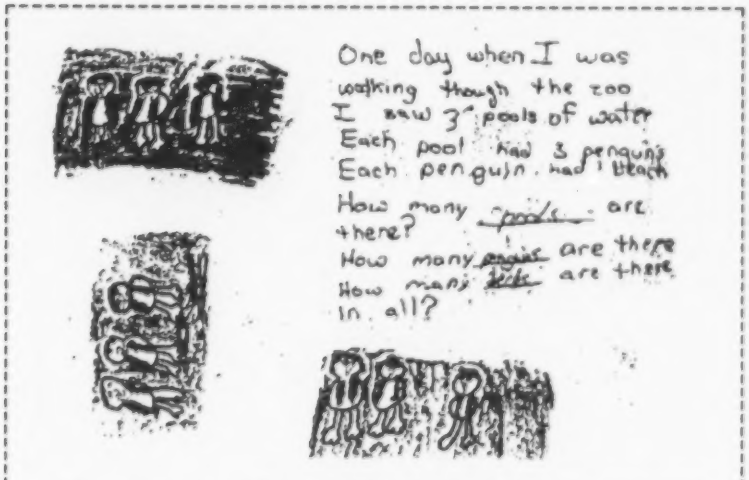
LANGUAGE ARTS

notes

SAMPLE STUDENT WORK 3

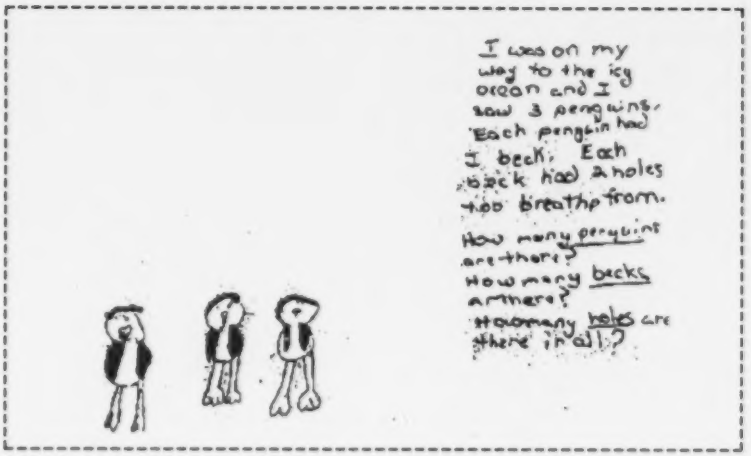

One day when I went walking I saw two black & white penguins On each penguin their was 2 eyes On each eye there was a dot
How many penguins are there?
How many eyes are there?
How many dots are there in all.

SAMPLE STUDENT WORK 3

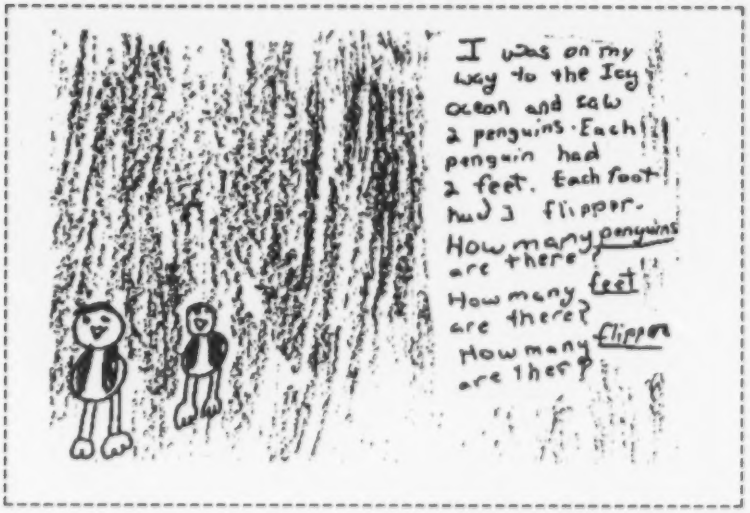

One day when I was walking though the zoo I saw 3 pools of water
Each pool had 3 penguins
Each penguin had 1 beach
How many pools are there?
How many penguins are there
How many beaks are there
in all?

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 3


I was on my way to the icy ocean and I saw 3 penguins. Each penguin had 1 beak. Each beak had 2 holes to breathe from.
How many penguins are there?
How many beaks are there?
How many holes are there in all?

SAMPLE STUDENT WORK 3


I was on my way to the Icy ocean and saw 2 penguins. Each penguin had 2 feet. Each foot had 3 flipper.
How many penguins are there
How many feet are there?
How many flipper are ther?

SCORE POINT 2

The graphic organizer is complete, but there is confusion about the relationships between the topics, subtopics, and examples. Some of the information indicated (e.g., little, big, small, and medium for height) is too vague to serve its purpose in the book. More than one Student Planning Sheet was completed, but as the accompanying example indicates, there is confusion about what is important and what is not. The book is complete, but the pattern is not consistently followed and the illustrations do not always match the text. There are also some inaccuracies in the information taken on the note taking sheets and presented in the book.

notes

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 4

Name: _____

STUDENT PAGE C - PLANNING SHEET

Topic: insect Subtopic: Velvet mite

Fact: these creatures are red and so small that you can barely see them.

Related Information: About thirty of them could fit on the fingernail of your thumb. velvet mite has 8 legs

A Number/Item	B Number/Characteristic or Association for A	C Number/Characteristic or Association for B
3 fingernails	1 velvet mite	8 legs

Setting: when I was walking to my grandmoms

Narrative: when I was walking to my grandmoms I noticed 3 of my fingernails. under each fingernail there was a velvet mite. each velvet mite had 8 legs.

Questions: How many fingernails are there? How many velvet mites there?

How many legs are there in all?

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 4

Name: _____

STUDENT PAGE C - PLANNING SHEET

Topic: insect Subtopic: grasshoppers

Fact: Grasshoppers can jump really well. If you try to catch one, it will usually jump away just as you are about to get it.

Related Information: Has insect with wings and strong back legs for jumping

A Number/Item	B Number/Characteristic or Association for A	C Number/Characteristic or Association for B
18 insects	2 legs	2 wings

Setting: when I was mowing the lawn

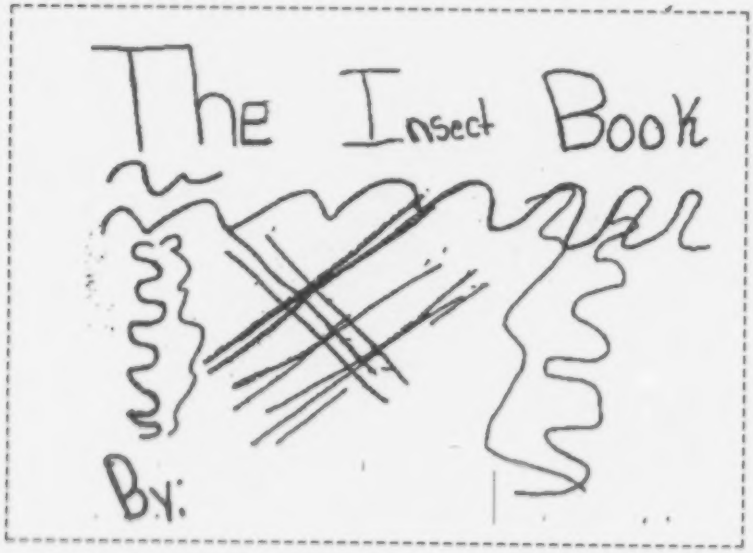
Narrative: when I was mowing the lawn it was full of grasshoppers. each grasshopper had 2 legs. Each leg had 2 wings.

Questions: How many grasshoppers are there? How many legs are there

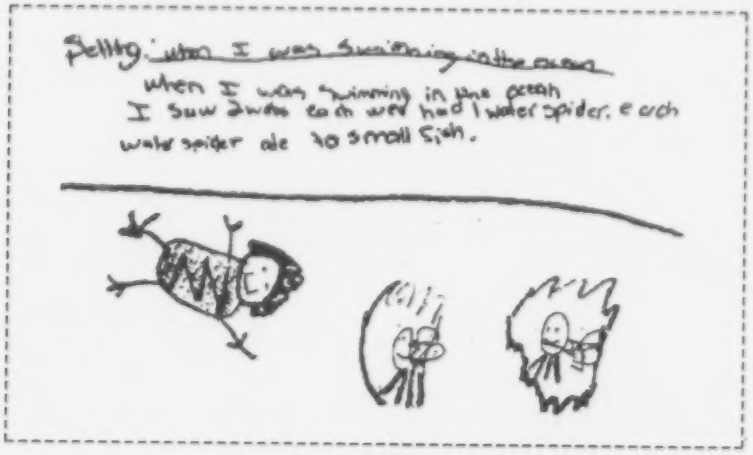
How many wings are there in all?

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 4



SAMPLE STUDENT WORK 4


Spelling: when I was Swimming in the ocean
When I was swimming in the ocean
I saw a water spider. Each
water spider ate 10 small fish.

BEST COPY AVAILABLE

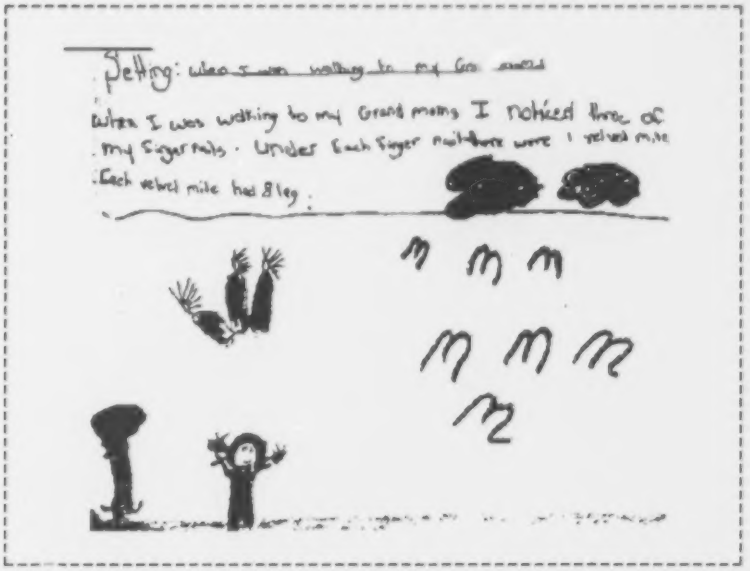
notes

notes

SAMPLE STUDENT WORK 4

Spelling: when I was mowing the lawn
When I was moing the lawn I saw 5 grasshoppers.
Each grasshopper had 2 legs. Each leg had 2 veins

SAMPLE STUDENT WORK 1


Spelling: when I was walking to my Grandmoms
When I was walking to my Grandmoms I noticed three of
my fingernails. Under Each fingernail there were 1 velvet mite
Each velvet mite had 8 leg.

BEST COPY AVAILABLE

SAMPLE STUDENT WORK 4

Setting
One day when I was walking to the store I saw 3 Ants. Each one had three. Each head had 3 antlers.
How many ants are there? How many heads are there?
How many antlers are there in all?
Store
Ants

SCORE POINT 1

No graphic organizer was submitted. More than one Student Planning Sheet was turned in, but much of the information was irrelevant and some of it was inaccurate. The book is complete and the illustrations make an attempt to match the text; however, the pattern of the narrative with a series of questions is not followed until the last page.

notes

BEST COPY AVAILABLE

notes

Developed by Bernie Floriani, The Delaware Department of Public Instruction, and Douglas Grudzina, Dover High School.

State Content Standard

Use oral and written English appropriate for a variety of purposes and audiences.

State Performance Standard

Student will write a persuasive letter or editorial piece that appropriately addresses the audience, demonstrates clarity and logic, and indicates control of grammar and mechanics.

Performance Task

Student will write a letter to parents, teacher, or other adult to request something. In the letter, student clearly states what he or she wants and provides good reasons for the suggestion.

Sample Lesson (Grades 2-6)

In order to facilitate the development of persuasive writing skills, teachers need to provide children with concrete opportunities to acquire a cognitive grasp of the higher order thinking processes that they are being asked to apply. The "Think Bag" teaching/learning activity stimulates student interest, elicits multiple examples, and sets the stage for the introduction of the persuasive writing performance task or prompt.

INITIATING ACTIVITY

Put the following items into a brown paper bag:

- ☑ A crumpled note in child's print (you can write it) that reads, "Dear Mom, I know what you said, but I want him to go with us. He will not like it at that place. I know. Love, Billy."
- ☑ A dog biscuit (or dog bone or toy).
- ☑ An adult note reading, "Ann, I'll be home by 5:00. Please have everything ready to go. Love, Bill."
- ☑ A handwritten list entitled: Things We Have to Take. Include in the list such items as hot dogs, catsup, mustard, three sleeping bags, ice chest, charcoal, matches, and tent.
- ☑ Any picture of a dog.
- ☑ A paper plate or cup.
- ☑ A picture from a magazine or a photograph of a place where people might camp.

87

notes

Challenge the children to examine the things in the bag. Have different students pull one item out of the bag at a time, discuss and show the item to the class, and infer each item's relationship to the others until all items have been removed from the Think Bag.

Encourage them to clarify their reasons for making particular inferences. Help them clarify their reasoning by encouraging them to examine and discuss the items among themselves.

The relationships among the items taken from the bag should lead children to infer that a family is going on a camping trip. Ask the following questions after this response is solicited:

"How many people are going?" (Three.) "How do you know?" (Three sleeping bags.) "What are their names?" (Ann, Billy, and Bill.) "Is the dog going on the trip?" "How effective is Billy's letter to Mom?" "Could it be rewritten to be more persuasive?" "What would need to be added, deleted, or changed?"

Have the students work in pairs, groups, or individually to outline their ideas in a graphic organizer (see page 79).

The use of the graphic organizer (page 79) helps students visualize their thoughts while encouraging them to put their initial ideas on paper so as not to forget them. This activity is not assessed. This type of thinking/planning is essential, however, as the ultimate quality of the students' persuasive letter will hinge on careful choice of an appropriate issue and audience.

Have students rewrite Billy's letter using the possible objections and possible arguments listed in the graphic organizer.

After the students have finished their drafts, provide several students with the opportunity to share their letters with the entire class. Then share the model student letter for them to compare to their own drafts.

Purpose:

WHO (AUDIENCE)	WHAT (SUBJECT)	WHY (REASONS)	POSSIBLE OBJECTIONS	POSSIBLE ARGUMENTS

notes

MODEL STUDENT LETTER

Dear Mom,

I would really like to have Waldo go with us on the camping trip next weekend. He will be lonely for us at the kennel, and they won't care for him as well as we would. I know you probably think that Waldo will make dog messes all around, but I could be a big help in cleaning them up. You might think that he will bark too much. You should know that I have gotten really good at quieting him down when he gets excited. I know that you are concerned about someone feeding him, but I promise to take full responsibility. I will keep food in his dish and water in his bowl. I hope you will think about my ideas and say, "yes." Please let me know what you decide.

Love,
Billy

Distribute to the students the scoring criteria (see pages 85-87 and 90-92) that will be used to assess their letters and have them work in groups scoring their own drafts, their peers', and the model. Encourage them to discuss specific elements of the various drafts and the scoring rubric to determine how they work toward creating effective persuasion.

Having students revise Billy's letter allows them to practice writing a persuasive letter, improving on one that they have already decided is flawed and allowing them to put into practice the necessary elements of persuasion that they have identified. Sharing their drafts with one another and comparing their drafts with the model will further emphasize what the elements of persuasion may look like in application. Sharing the scoring criteria with students and having them score their own, their peers', and the model revised "Billy letter" will also demonstrate how the elements of persuasion can be used both to develop and to evaluate the effectiveness of a persuasive piece.

BE CREATIVE

If the family camping trip scenario is not appropriate for your class, develop your own scenario and fill the "think bag" with the appropriate artifacts. Draft an appropriate model letter, making certain your model exemplifies the points that the scoring rubric specifies.

notes

POTENTIAL FOR INTEGRATION

Imagine the "think bag" as the purse, briefcase, or suitcase of a famous person — historical, current, or fictional. Fill the bag with artifacts that offer clues to this person's identity and draft an appropriate persuasive or informative letter. Or fill the bag with clues to help identify a civilization, culture, or society that the students have studied. Draft an appropriate letter.

PREWRITING

1. Ask the class to brainstorm *who* they have tried to persuade in the past, *what* they have tried to persuade them to do, *how* they tried to persuade them, and *what* the results were, and record their answers on the following chart.

WHAT	WHO	HOW	RESULTS
Take me to the movies	Big brother	Begged	Was mad but took me anyway
Let me take skating lessons	Mom	Pouted and whined	Said, "No."
Stay all night	Friend	Asked politely	Stayed
Buy me a bike	Parents	Cried	Didn't buy it

2. Ask students to describe and explain orally their subjects and audiences (what, who) to the class. Discuss the how and results columns.

3. Ask the class if anyone has ever tried writing a persuasive letter. If nobody has, suggest it and explain that letter writing can be a very effective tool for persuasion.

Prewriting activities generate ideas for writing. In any of its wide range of forms — class discussion, brainstorming, visualizing, free writing, and so on — prewriting aims to stimulate the free flow of thought. In this lesson, students are asked to share experiences in which they tried to persuade someone to do something, to allow something, or to give something. This activity stimulates student interest, elicits a wealth of examples, and sets the stage for introducing the prompt.

4. Introduce the prompt to the students.

notes

THE PROMPT

Think about something that you would like to persuade someone to do. It may be something you want someone to buy or do for you. Perhaps you would like someone to change his or her behavior or attitude towards you or someone you know. It could be that you would like to see a new rule made, enforced, or changed.

Who is the person you must persuade? What do you want to persuade him/her to do?

Write a letter in which you attempt to persuade this person to do what you want him/her to do. Your letter should show that you have:

- ✔ *Clearly stated what you want and why.*
- ✔ *Predicted at least two possible objections your audience might have.*
- ✔ *Shown logically that those objections have been considered and resolved.*

The prompt invites students to address the rhetorical specifications of subject, occasion, audience, and purpose. The graphic organizer likewise encourages students to plan to address these necessary rhetorical specifications.

5 Ask students to choose one thing they would like to persuade someone to do (who, what) and enter the information in the first two columns of their graphic organizers (see page 79). Students may work in pairs, in groups, or individually.

6 Have students enter their reasons (why) on the chart.

7 Have students enter the possible reactions of the audience and possible arguments of the persuader on the chart.

While addressing potential audience objections does not appear in the Performance Indicators for the content standard until grade 10, it is appropriate to have students begin to do this earlier to allow them time to master this persuasive technique and develop the habit of doing it.

8 Review the prompt and proper letter format with the class. Discuss possible opening statements that they could use in their letters. (It is helpful to list them on the chalkboard.)

notes

SAMPLE LETTER FORMAT

	(date)
Greeting:	*Dear Mary,*
Body:	*It is really important to me that . . .*
Closing paragraph:	*So, you see, this is why . . .*
Closing:	*Sincerely,*
	(Signature)

9 Read a model to the class emphasizing the structure followed. Students may use their own structure, but should include

- ✔ What is wanted
- ✔ Reasons for wanting it
- ✔ Two or three possible objections
- ✔ Reasons to overcome the objections
- ✔ Closing summary

The following is a sample of the model.

(date)

Dear Mom,

This year I would like to have my birthday party at Farrell's with three of my best friends. I've always wanted to go there because they sing "Happy Birthday" and play the big drum if you tell them it's your birthday.

I know you probably will think that it will be too expensive, but it really won't be because I will pay for my friends' ice cream with my allowance. You won't need to give me any extra money because my ice cream will be free just because it's my birthday. That's why everyone likes to go to Farrell's on their birthday.

You might not know where there is a Farrell's and be worried about driving with kids in the car. Guess what? There is a Farrell's just two blocks from school. We could walk and meet you there.

I hope you will think about my idea and say "Yes." The only thing I really want for my birthday is to have a party at Farrell's. Please let me know what you decide.

Love,
Molly

93

notes

WRITING

1. Students write their letters referring to their completed graphic organizers.

2. Sharing: Students share letters and help each other decide whether or not their letters will persuade the intended audiences. Partners should underline the objections of the audience in red and arguments to overcome objections in green. Partners should also discuss whether or not the language in each letter is appropriate for its audience.

3. Revising: Based on feedback received, students should revise their letters to make them more persuasive. They should consider the following questions:

   - ✔ Is what I want clearly stated?
   - ✔ Have I stated the reasons that I want it?
   - ✔ Did I include possible objections?
   - ✔ Are my arguments against those objections persuasive?
   - ✔ Is my closing paragraph effective?

4. Editing: Students may edit their own letters (or exchange them with a friend).

5. Submission and publication. Collect all final drafts and score using the scoring criteria already shared with students (see pages 85-87 and 90-92).

Scoring Guide (Grade 2)

notes

5 POINTS

- ☑ The writer's position on the issue is clear and unambiguous.
- ☑ Reasons, opinions, and examples offered are logical and show an awareness of the intended audience.
- ☑ There is a sense of sufficiency and completeness in the reasons, opinions, and examples offered.
- ☑ The letter shows evidence of thought and clear reasoning.
- ☑ The letter has a clear and recognizable progression of ideas.
- ☑ The words are precise, vivid, and economical.
- ☑ Sentences are complete and varied in length and structure.
- ☑ Appropriate grammar, mechanics, spelling, and usage enhance the meaning and readability of the text.

4 POINTS

- ☑ The writer's position on the issue is clear.
- ☑ Reasons, opinions, and examples offered are for the most part logical and show some awareness of the intended audience.
- ☑ There is a sense of sufficiency and completeness in the reasons, opinions, and examples offered.
- ☑ The letter shows some evidence of thought and reasoning.
- ☑ The letter has a clear and recognizable progression of ideas.
- ☑ The words are precise, vivid, and economical.
- ☑ Sentences are complete and varied in length and structure.
- ☑ Appropriate grammar, mechanics, spelling, and usage enhance the meaning and readability of the text.

notes

3 POINTS

- ✔ The writer's position on the issue is unclear or ambiguous.
- ✔ Reasons, opinions, and examples offered seem illogical or irrelevant and show little awareness of the intended audience.
- ✔ There is a sense of insufficiency or incompleteness in the reasons, opinions, and examples offered; the letter leaves the reader wanting "more."
- ✔ The letter shows little evidence of thought and reasoning; it is somewhat difficult for the reader to follow the writer's point.
- ✔ Word choice is limited.
- ✔ Sentences, while generally complete, are limited in variety of length and structure; there may be some errors in sentence formation, but these do not hinder the reader's understanding of the letter.
- ✔ Problems in grammar, mechanics, spelling, and usage may appear, but they do not seriously affect the meaning and readability of the text.

2 POINTS

- ✔ The writer's position on the issue is unclear or ambiguous.
- ✔ Reasons, opinions, and examples offered are illogical or irrelevant and show little awareness of the intended audience.
- ✔ Reasons, opinions, and examples offered are insufficient; the letter may state and repeat only one reason or example.
- ✔ The letter shows very little evidence of thought and reasoning; it is difficult for the reader to follow the writer's point.
- ✔ Word choice is limited, inaccurate, or inappropriate.
- ✔ Sentences are limited in variety of length and structure, and there are errors in sentence formation that begin to hinder the reader's understanding of the letter.
- ✔ Problems in grammar, mechanics, spelling, and usage begin to affect the meaning and readability of the text.

WRITING STANDARDS

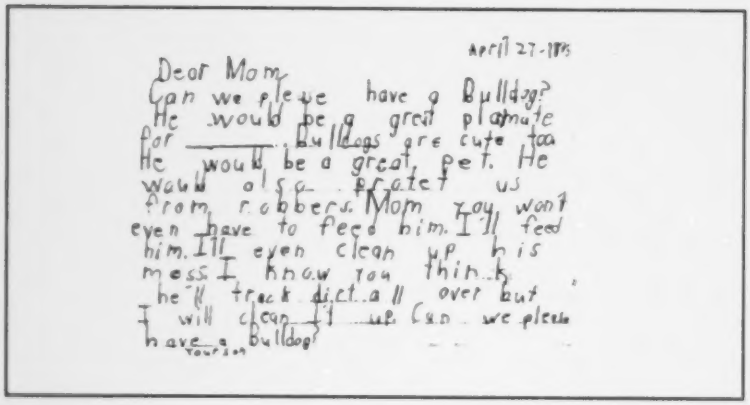
1 POINT

- ☑ The writer's position on the issue is unclear or ambiguous.
- ☑ No reasons, opinions, or examples are offered in support of the position.
- ☑ The letter shows no evidence of thought and reasoning; it is very difficult or impossible for the reader to follow the writer's point.
- ☑ Word choice is limited, inaccurate, or inappropriate.
- ☑ Errors in sentence formation hinder the reader's understanding of the letter.
- ☑ Problems in grammar, mechanics, spelling, and usage severely affect the meaning and readability of the text.

notes

notes

SAMPLE STUDENT LETTER 1 — (GRADE 2)



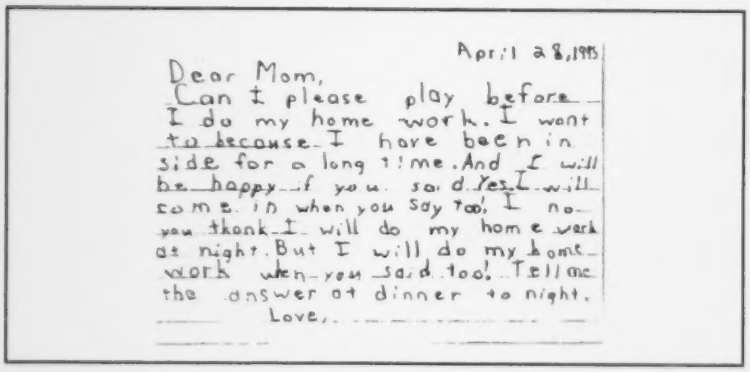
April 27-1995

Dear Mom,
Can we pleye have a Bulldog?
He would be a great playmate
for _____. Bulldogs are cute too
He would be a great pet. He
would also protect us
from robbers. Mom you won't
even have to feed him. I'll feed
him. I'll even clean up his
mess. I know you think
he'll track dirt all over but
I will clean it up. Can we pleze
have a bulldog?
Yours,

SCORE POINT 5

This letter states a clear and unambiguous position ("Can we have a bulldog?"). The writer follows the initial request with several reasons (great playmate for brother, bulldogs are cute, great pet, and protection from robbers). The writer also directly addresses the needs of his audience (his mother) by stating that he will feed the dog and clean up his mess. There are no severe errors in the conventions of standard written English to render this letter difficult to read.

SAMPLE STUDENT LETTER 2 — (GRADE 2)



April 28,1995

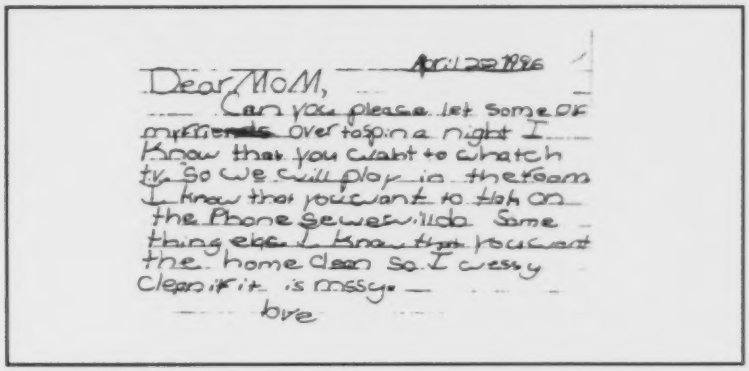
Dear Mom,
Can I please play before
I do my home work. I want
to because I have been in
side for a long time. And I will
be happy if you said Yes I will
come in when you say too. I no
you thank I will do my home work
at night. But I will do my home
work when you said too. Tell me
the answer at dinner to night.
Love,

SCORE POINT 4

While this letter has the content of a 5-point letter (clear position and ample support), errors in spelling and punctuation begin to affect its readability. The initial question lacks the appropriate ending punctuation. Misspellings like "too" for "to" and "thank" for "think" make the reader pause to figure out what the writer really means. This letter could probably be cleaned up in another draft and would receive a higher score.

BEST COPY AVAILABLE

SAMPLE STUDENT LETTER 3 — (GRADE 2)



Dear MoM, April 28 1996

Can you please let some of
~~my friends~~ over to spin a night I
know that you wabt to watch
tv. So we will play in the foam
I know that you want to talk on
the Phone Sewe will do Some
thing else I know that you want
the home clean so I wessy
Clean it it is messy.

love

SCORE POINT 2

While this letter has a clear position (let some friends spend the night), there are no actual reasons cited as to why the writer wants his or her friends to come or why the mother should allow it. There are also some fairly serious errors in standard written English that render this letter difficult to read. The writer promises to play in the "foom" so the mother can watch television and he/she promises "wessy" will clean the house if it gets messy.

notes

BEST COPY AVAILABLE

Scoring Guide (Grade 5)

notes

5 POINTS

- ☑ The writer's position on the issue is clear and unambiguous.
- ☑ Reasons, opinions, and examples offered are logical and show an awareness of the intended audience.
- ☑ Support from experts is accurate and relevant.
- ☑ There is a sense of sufficiency and completeness in the support of the position.
- ☑ The letter shows strong evidence of thought and clear reasoning.
- ☑ The letter has a clear and recognizable progression of ideas.
- ☑ The words are precise, vivid, and economical.
- ☑ Sentences are complete and varied in length and structure.
- ☑ Appropriate grammar, mechanics, spelling, and usage enhance the meaning and readability of the text.

4 POINTS

- ☑ The writer's position on the issue is clear.
- ☑ Reasons, opinions, and examples offered are for the most part logical and show some awareness of the intended audience.
- ☑ Support from experts is for the most part accurate and relevant.
- ☑ There is a sense of sufficiency and completeness in the support of the position.
- ☑ The letter shows some evidence of thought and reasoning.
- ☑ The letter has a clear and recognizable progression of ideas.
- ☑ The words are precise, vivid, and economical.
- ☑ Sentences are complete and varied in length and structure.
- ☑ Appropriate grammar, mechanics, spelling, and usage enhance the meaning and readability of the text.

notes

3 POINTS

- ☑ The writer's position on the issue is unclear or ambiguous.
- ☑ Reasons, opinions, and examples offered seem illogical or irrelevant and show little awareness of the intended audience.
- ☑ Support from experts may be partially inaccurate or irrelevant.
- ☑ There is a sense of insufficiency or incompleteness in the support of the position; the letter leaves the reader wanting "more."
- ☑ The letter shows little evidence of thought and reasoning or the reasoning is flawed; it is somewhat difficult for the reader to follow the writer's point.
- ☑ Word choice is limited.
- ☑ Sentences, while generally complete, are limited in variety of length and structure; there may be some errors in sentence formation, but these do not hinder the reader's understanding of the letter.
- ☑ Problems in grammar, mechanics, spelling, and usage may appear, but they do not seriously affect the meaning and readability of the text.

2 POINTS

- ☑ The writer's position on the issue is unclear or ambiguous.
- ☑ Reasons, opinions, and examples offered are illogical or irrelevant and show little awareness of the intended audience.
- ☑ Support from experts is inaccurate or irrelevant.
- ☑ Support for the position is insufficient; the letter may state and repeat only one reason or example.
- ☑ The letter shows very little evidence of thought and reasoning, or reasoning is flawed and does not seem to make sense; it is difficult for the reader to follow the writer's point.
- ☑ Word choice is limited, inaccurate, or inappropriate.
- ☑ Sentences are limited in variety of length and structure, and there are errors in sentence formation that begin to hinder the reader's understanding of the letter.
- ☑ Problems in grammar, mechanics, spelling, and usage begin to affect the meaning and readability of the text.

notes

1 POINT

- ☑ The writer's position on the issue is unclear or ambiguous.
- ☑ No reasons, opinions, or examples are offered in support of the position.
- ☑ The letter shows no evidence of thought and reasoning; it is very difficult or impossible for the reader to follow the writer's point.
- ☑ Word choice is limited, inaccurate, or inappropriate.
- ☑ Errors in sentence formation hinder the reader's understanding of the letter.
- ☑ Problems in grammar, mechanics, spelling, and usage severely affect the meaning and readability of the text.

SAMPLE STUDENT LETTER I — (GRADE 5)

Purpose: To have no homework for the rest of the year.

Who (Audience)	What (Subject)	Why (Reasons)	Possible Objections	Possible Arguments
Mrs. _____	No Homework	- It's to nice outside to be stuck inside doing homework. - We have sports and other activity's at night and in the afternoon that don't give us a chance to do homework. - We'll have a lot of homework next year so why not give us a break for the rest of this year.	- It's not the end of the year yet, you still need to learn. - Doing homework for the rest of the year will help you get ready for next year or Redding. - You have all summer to play so just wait until then.	- I think that we learn and cover enough in school that we don't need homework. - We have done enough in school that has already got me ready for Redding. Plus after school instead of homework we could be doing things that we wouldn't have a chance to do in Redding, like playing outside and doing sports. - But people might be doing something that won't allow them to get outside and play.

5-23-95

Dear Mrs. _____,

Could you please think about not giving us anymore homework for the rest of the year? I ask this because it's too nice outside to be stuck inside doing loads of homework. Besides, some of us have sports and other activities in the afternoon which doesn't give us a chance to do our homework. But in some cases sports let us learn too, so it wouldn't be like we would stop thinking. Also, we'll have tons of homework next year, so why not give us the rest of the year off.

I know you might say that the years not over and we still need to learn and homework will get us ready for next year at _____, but I think we have learned and covered enough already. I also know that we will still continue moving work which will sharpen our skills.

I think you also might say that we have all summer to play, but some people might be doing something that won't allow them to get out.

Please just think about it.

Sincerely,

P.S. You don't really like correcting homework, do you?

BEST COPY AVAILABLE

SCORE POINT 5

This very strong letter has a clear and unambiguous point (no more homework) and several thoughtful reasons. The writer has clearly anticipated the teacher's objections and offers reasonable responses. Finally, the letter is virtually free of errors in standard written English.

SAMPLE STUDENT LETTER 2 — (GRADE 5)

Purpose: To convince my Mom & Dad to let me try out for the football team.

Who (Audience)	What (Subject)	Why (Reasons)	Mom & Dad Possible Objections	Me Possible Arguments
Mom & Dad	To convince Mom & Dad to let me try out for the football team	•I like to play football •I love to run and you run a lot in football •I know all the rules so I will play fair.	•You might get hurt. •I'll have to drive you to all your games and practices •It costs to much money •You probably won't play much •The coaches are strict •You might not even make the team.	•We wear pads and helmets •Mayle Mom Mom can drive me to practices. •I'll pay half. •I going to work hard so he will play me •so are my soccer coaches •I'm going to try my hardest

Dear Mom & Dad, steps

I know you said no about trying out for the football team but I've been practicing a lot. I love football.

I know its a lot of money to try out but I've been saving my allowance I understand that transportation is a problem but I can ask Mom Mom or Pop Pop to take me to some practices. I know you're afraid that I'll get hurt, but we wear pads and helmets. The coaches are strict but I'm used to it because my soccer coaches are strict too. I know you think I won't play much because I'm small, but everything I do I'm good at.

Can you please think about it for a while? Thanks.

Your Son.

BEST COPY AVAILABLE

SCORE POINT 4

This letter has many of the necessary elements of a 5-point letter, but the writer does not state clear reasons that his parents should allow him to play football. There is an awareness of audience in most of the objections cited, but the responses are not as strong or compelling as they might be (e.g., "I can ask Mom Mom or Pop Pop" rather than stating that they'll be able to help; "everything I do I'm good at" rather than a more thoughtful response to "I won't play much").

SAMPLE STUDENT LETTER 3 — (GRADE 5)

Purpose: To persuade parents on going to the mall

Who (Audience)	What (Subject)	Why (Reasons)	Possible Objections	Possible Arguments
parents	Persuading parents on going to the mall.	-Haven't been there in a while -need new clothes or shoes -anyway might need a work something -could pay for it with our money	But can go the mall might - can always get new clothes or shoes - you can always get what you want and some place else - Costs a lot -it all?	But there is a sale on anything and today is the last day They might not have the right size -But no other place sells what you want and it's a limited time only - found out it clothes cost too much

Dear Mom, May 19, 1995

Could we please go to the mall? We haven't been there for a long time. I understand you haven't felt like going to mall lately but you said that you felt better. Anyway I need new clothes and shoes because I have just about outgrown my shoes, and I'm starting to outgrow my clothes too. I really do need them. I understand that you think I really don't need new shoes because it don't look as if I need them, but I really do. Besides I heard you and dad both want something from the mall. Anyway I could pay for the clothes and shoes with my own money. Again I understand that think it will cost a lot of money, but I found out that clothes and shoes are on sale. Come on mom. I promise I'll do all of my chores and get my homework done early. Please? Sincerly,

BEST COPY AVAILABLE

SCORE POINT 3

This "low 3" has some potential, but a number of problems prevent it from receiving a higher score. The point is ambiguous, starting out as an appeal to go to the mall, but becoming by the end of the letter an argument in favor of buying new clothes. While there are a few reasons cited (outgrowing clothes and shoes), again the writer seems distracted by the issue of shoes ("it don't look as if I need them, but I really do"). The response to the parents' objection ("I really do") shows little thought about what will convince the parents. Similarly, the reason, "I heard you and dad both want something from the mall," is oversimplified and not completely thought out. By the end of the letter, the writer has completely abandoned a logical or reasonable appeal and offers "I promise I'll do all of my chores and get my homework done early." Finally, there are severe errors in standard written English ("it don't look," "you haven't felt like going to mall") that need to be corrected for a higher score.

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

INTRODUCTION

IMPLEMENTATION MODELS AND PROTOTYPES 3

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

POLICY

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

ANALYSIS OF STATE CONTENT STANDARDS

The following analyses of standards in reading, writing, mathematics, science, and social studies should prove helpful to states in determining how well they have covered the content in their own standards. The analyses provide information on standard specificity, formatting of documents, and characteristics of appropriate explanations and examples.

There are several ways the analyses can assist state or district efforts:

- ✔ States or districts that are beginning to develop standards can review these analyses to gather information on what has been done and to determine what standards from other states may be potential models and resources.
- ✔ States or districts can use these analyses to review and consider the completeness of their own standards.
- ✔ States can use these analyses as an initial step in determining the possibility of interstate collaborations for developing assessments, accountability tools, curriculum materials, and staff resources.

The level of the analyses varies. This is a consequence of the differences in the character and organization of the content domains. However, some of the common characteristics of the standards across the domains are also discernible.

PAGE 2

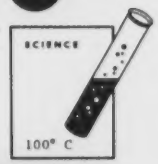
PAGE 7

READING — PAGE 12
WRITING — PAGE 14
ENGLISH LANGUAGE ARTS — PAGE 15

PAGE 27

BEST COPY AVAILABLE

Comparability of State Science Standards Documents (Standards completed since 1993)

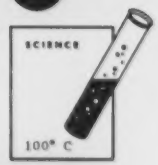


	SCIENCE AS INQUIRY	PHYSICAL SCIENCE	LIFE SCIENCE	EARTH AND SPACE SCIENCE	SCIENCE AND TECHNOLOGY	SOCIAL/ ENVIRONMENTAL	HISTORY/NATURE OF SCIENCE	UNIFYING CONCEPTS/ PROC.	OTHERS
CO	✔	✔	✔	✔	✔				✔
DE	✔	✔	✔	✔	✔	✔	✔		
DC		✔	✔	✔				✔	
FL	✔	✔	✔	✔		✔	✔	✔	✔
KS	✔	✔	✔			✔	✔	✔	
KY	✔				✔			✔	✔
MA	✔	integrated			✔	✔	✔	✔	
NC		✔	✔	✔			✔	✔	✔
ND	✔	✔	✔					✔	
NE		✔	✔	✔				✔	✔
NH	✔	✔	✔	✔	✔			✔	
NY	✔	✔	✔		✔	✔		✔	✔
OH	✔	✔	✔		✔	✔	✔	✔	✔
OK	✔	✔	✔	✔					✔
UT	✔	✔	✔	✔			✔	✔	

Note: Standards categories based on NRC *Science Education Standards*, 1995 and AAAS *Benchmarks*, 1993. Council of Chief State School Officers, State Education Assessment Center, Washington, DC, 1996.

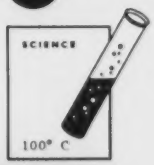
BEST COPY AVAILABLE

Descriptive Characteristics — Science Standards



	NUMBER OF CONTENT STANDARDS (K-12)	GRADE LEVEL BENCHMARKS	COMBINED SCIENCE AND MATH	COMMUNICATING CONTENT STANDARDS				CONTENT STANDARDS AND POLICIES OR PLANS FOR:			
				VIGNETTES	EXAMPLES/ MODELS	ACTIVITIES/ TASKS	DIAGRAMS/ GRAPHICS	PROFESSIONAL DEVELOPMENT	STATE ASSESSMENT	MATERIALS AND TEXTS	TECHNOLOGY INTEGRATION
CO	8	8 (K-4, 5-8, 9-12)	✔						L		
DC		8 (K-4, 5-8, 9-12)	✔			✔		R	R	R	R
DE	8	40-50 (K-4, 5-8, 9-12)			✔	✔	✔	R	R	R	R
FL	8	8 (K-4, 5-8, 9-12)		✔	✔	✔	✔	R	R		
KS	4	8-16 (2, 5, 8, 11)				✔					
KY	6		✔	✔	✔	✔	✔	L	L	L	L
MA	4	60-100 (K-4, 5-8, 9-12)		✔	✔	✔	✔	L	R	R	R
NC		6-10 per grade					✔				
ND	6		✔								
NE		20-30 (K-4, 5-8, 9-12)	✔	✔		✔	✔				
NH	6	30 (6, 10)									
NY	9		✔		✔	✔	✔	R	R		
OH	5	50 per grade		✔	✔	✔	✔		L	R	R

BEST COPY AVAILABLE



Descriptive Characteristics — Science Standards (continued)

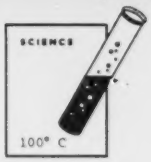
	NUMBER OF CONTENT STANDARDS (K-12)	GRADE LEVEL BENCHMARKS	COMBINED MATH AND SCIENCE	COMMUNICATING CONTENT STANDARDS				CONTENT STANDARDS AND POLICIES OR PLANS FOR:			
				VIGNETTES	EXAMPLES/ MODELS	ACTIVITIES/ TASKS	DIAGRAMS/ GRAPHICS	PROFESSIONAL DEVELOPMENT	STATE ASSESSMENT	MATERIALS AND TEXTS	TECHNOLOGY INTEGRATION
OK		6-12 per grade	✔								
PA		8 (3, 6, 9, 12)				✔					
UT		5-15 per grade/course				✔			L		

✔ = Method used. R = Framework recommends linkage. L = Existing linkage reported in framework.

Comparing the Terminology of Science Standards Documents

STATE	CONTENT STANDARD
FL (1994)	The Florida framework describes examples of concepts and activities by grade block within 8 broad knowledge strands of what students should be able to know and do. For example, Strand 6: Processes of Life: (1) Design a model to demonstrate the correlation between healthful living and human body-system maintenance.
IN (1993)	Indiana defines 10 science proficiencies and 30 content themes (K-12). For example, Theme H: Diversity of Life: Students will understand and use the concept that living organisms, though they share many similarities, are richly diverse in structure and function.
KS (1993)	Content is organized along 4 broad program outcomes; each indicates 2 to 6 learner goals (K-12) and benchmarks for grades 2, 5, 8, and 11. For example, Grade 8: Understands that forms and interactions of matter and energy determine the nature of the environment.
MA (1994)	Massachusetts defines fundamental understandings and skills under science content areas and domains with 60 skills defined for grades K-4, 84 for grades 5-8, and 103 for grades 9-12. For example, Life Sciences: Diversity and Adaptation of Organism: short term changes in available food, moisture, or temperature of an ecosystem may result in a change in the number of organisms in a population.
NC (1994)	North Carolina defines competency goals and objectives by grade and course with a total of 65 for grades K-12. For example, Grade 6, Competency Goal 6: The learner will construct an understanding of science concepts through patterns and cycles in the natural world. Objective 6.1: Investigate similarities and differences among monerans, protist, fungi, plants, and animals.
ND (1993)	North Dakota defines content outcomes with 6 in science. For example, by the end of grade 8, students develop and use tables, graphs, and rules to identify properties and relationships.
NE (1994)	Nebraska defines measurable performances by strand, with 19-27 per 4 science levels. For example, Science Strand: Cells and heredity: Design an investigation demonstrating how materials move in and out of cells.
NH (1994)	New Hampshire defines 6 curriculum strands with curriculum standards for grades K-12, and proficiency standards for grades 6 and 10. For example, Curriculum Standard: Students will demonstrate an increasing ability to recognize patterns and products of evolution, including genetic variation, specialization, adaptation, and natural selection.

BEST COPY AVAILABLE

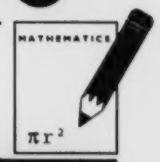


Comparing the Terminology of Science Standards Documents (continued)

STATE	CONTENT STANDARD
NV (1994)	Nevada defines specific learning objectives (content and process) by grades 3, 6, and 8, and for categories of high school courses. For example, By completion of grade 8 the student will: demonstrate understanding of these concepts and principles: 1. Diversity: natural phenomena display a wide variety of similarities and differences.
NY (1994)	New York presents 9 content standards for math, science, and technology education, with each describing major areas of study, key competencies and concepts, and performance indicators. For example, Standard 4: Science, Living Environment: Describe the properties that distinguish living from nonliving things.
OH (1994)	Ohio provides sample grade level (instructional) objectives in 4 strands of scientific knowledge, with about 50 objectives per grade. For example, Strand, Scientific knowledge: Objective, Grade 7: The learner will investigate the interdependence and similarities of organizations and their physical environments (e.g., mimicry, camouflage).
OK (1993)	Oklahoma defines 6 to 12 student skills for science by grade under process and content categories. For example, Science Grade 8: 1. Observing and Measuring: The student will: A. Identify similar or different characteristics in a given set of objects, organisms, or events.
PA (1995)	Science and technology are benchmarked in four levels with transitional standards to address eight general state learning outcomes. Content knowledge, process skills, and curriculum integration are developed understanding of the characteristic within these contexts.
UT (1994)	Utah defines standards by science topic with 3 to 5 each for grades K-6, and 10 to 15 per grades 7-12 courses. For example, Life Science, Standard: The students will demonstrate an understanding of the characteristic processes of life. Objective (example): Describe the processes characteristic of all life: reproduction, movement, assimilation, growth, response, and energy transformation.

BEST COPY AVAILABLE

Comparability of State Mathematics Standards Documents (Standards completed since 1993)



	NUMBER SENSE/ OPERATIONS	MEASUREMENT	GEOMETRY	ALGEBRA/ FUNCTIONS	STATISTICS/ PROBABILITY DATA	PROBLEM SOLVING	COMMUNICATIONS	REASONING/ THINKING SKILLS	CONNECTIONS	OTHERS
AR	✔	✔	✔	✔	✔	✔	✔		✔	✔
CA	✔	✔	✔	✔	✔	✔	✔	✔	✔	
CO	✔	✔	✔	✔	✔	✔	✔	✔		✔
DC	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
DE	✔	✔	✔	✔	✔	✔	✔	✔	✔	
ID	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
KS	✔		✔	✔	✔		✔	✔		✔
KY	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
LA	✔	✔	✔	✔	✔					
MA	✔	✔	✔	✔	✔	✔	✔	✔	✔	
MS	✔	✔	✔		✔	✔				✔
ND	✔		✔	✔	✔					✔
NE	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
NH	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
NJ	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
NY	✔	✔	✔	✔	✔	✔				✔
OK	✔	✔	✔	✔	✔	✔	✔	✔	✔	
SC	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

Note: Standards categories based on NCTM *Curriculum & Evaluation Standards*, 1989. Council of Chief State School Officers, State Education Assessment Center, Washington, DC, 1996

BEST COPY AVAILABLE

Descriptive Characteristics — Mathematics Standards

	NUMBER OF CONTENT STANDARDS (K-12)	GRADE LEVEL BENCHMARKS	COMBINED MATH AND SCIENCE	COMMUNICATING CONTENT STANDARDS				CONTENT STANDARDS AND POLICIES OR PLANS FOR:			
				VIGNETTES	EXAMPLES/ MODELS	ACTIVITIES/ TASKS	DIAGRAMS/ GRAPHICS	PROFESSIONAL DEVELOPMENT	STATE ASSESSMENT	MATERIALS AND TEXTS	TECHNOLOGY INTEGRATION
AR	13				✔			R	L		
CA	14	14 (K-4, 5-8, 9-12)			✔	✔	✔	L	R	L	
CO	9	20-30 (K-4, 5-8, 9-12)	✔						L		
DC		30-35 (K-4, 5-8, 9-12)	✔			✔		R	R	R	R
DE	10	60 (K-3, 4-5, 6-8, 9-10)		✔	✔	✔	✔	R	R	R	R
ID		13 (K-4, 5-8, 9-12)			✔		✔	R	L		R
KS		15 (4, 8, 12)			✔	✔	✔		L		
KY	12		✔	✔	✔	✔	✔	L	L	L	L
LA	5										
MA	9	9 (K-4, 5-8, 9-12)		✔	✔	✔	✔	R	R	R	R
MS		20-30 per grade (K-8)									
ND	15		✔								
NE		20-30 (K-4, 5-8, 9-12)	✔	✔		✔	✔				

BEST COPY AVAILABLE

Descriptive Characteristics — Mathematics Standards (continued)

	NUMBER OF CONTENT STANDARDS (K-12)	GRADE LEVEL BENCHMARKS	COMBINED MATH AND SCIENCE	COMMUNICATING CONTENT STANDARDS				CONTENT STANDARDS AND POLICIES OR PLANS FOR:			
				VIGNETTES	EXAMPLES/ MODELS	ACTIVITIES/ TASKS	DIAGRAMS/ GRAPHICS	PROFESSIONAL DEVELOPMENT	STATE ASSESSMENT	MATERIALS AND TEXTS	TECHNOLOGY INTEGRATION
NH	8	40 (6, 10)									
NJ	18			✔	✔			R	R		L
NY	9		✔		✔	✔	✔	R	R		
OK		6-12 per grade	✔						L		L
SC	6	30-50 (K-3, 3-6, 6-9, 9-12)			✔	✔	✔	R	R	R	R
UT		13-16 per grade/course									

✔ = Method used. R = Framework recommends linkage. L = Existing linkage reported in framework.

BEST COPY AVAILABLE

Comparing the Terminology of Mathematics Standards Documents

STATE	CONTENT STANDARD
AR (1993)	Arkansas framework is based on 13 content standards applicable to all grades. For example, Content Standard 1 (algebra and functions strand): Student will use the language of algebra as a representation tool.
CA (1992)	California defines 13-14 content standards by grade block. For example, Standard 9: Algebra: In grades five through eight, the mathematics curriculum should include explorations of algebraic concepts and processes so that students can: understand the concept of variable, expression, and equation, represent situations and number patterns with tables, graphs, verbal rules, and equations.
CO (1994)	Colorado defines 9 content standards in math. For example, Math Standard 2: Students use algebraic methods to explore, model, and describe patterns and functions. In grades 5-8 students extend their knowledge to analyze functional relationships to explain how a change in one quantity results in a change in another.
DC (1993)	The District defines level outcomes for 7 math content strands and 3 science strands, with 30-35 outcomes, averaging 6-8 on each of 4 levels. For example, Mathematics: Algebraic Concepts and Operations: Apply the concepts of variable, expression, and equation. Represent situations and number patterns with tables, graphs.
ID (1994)	Idaho defines standards and performance objectives by grade block. Standards: 13 for grades K-4 , 13 for 5-8, and 14 for 9-12. For example, Standard IX: Algebra: All students will explore algebraic concepts and processes. Performance Objectives: All students will (1) understand the concepts of variable, expressions, and equation, (2) represent situations and number patterns with tables, graphs.
KS (1993)	Kansas defines curriculum outcomes with 15 outcomes for each of the 3 benchmark grade levels (4, 8, and 10). For example, Focus Area: Patterns, Functions, Mathematical/Algebraic Functions, Algebraic Concepts and Relationships; Outcome 6: recognizes and describes linear functions and uses appropriate graphing technology to graph and analyze linear functions.
KY (1993)	Kentucky defines learner outcomes with 12 math outcomes across the grade levels. For example, Math: Learner Outcome 2.8: Students demonstrate understanding of concepts related to mathematical procedures. Solve equations with concrete, pictoral, and abstract methods.
LA (1994)	Louisiana defines 5 interwoven themes that encompass the strands of mathematics. For example, Algebra, Grades 5-8: In problem-solving investigations, use appropriate manipulatives, models, graphs, tables, and technology to develop the understanding of concepts and to explore the applications of algebra. Explore student- and teacher-generated methods for solving equations and inequalities.

BEST COPY AVAILABLE

Comparing the Terminology of Mathematics Standards Documents (continued)

STATE	CONTENT STANDARD
MA (1994)	Massachusetts defines content standards by grade block, with 9 standards for grades K-4, 9 for 5-8, 10 for 9-12, and 7 for Adult Basic Education. For example, Grades 5-8, Standard 9: Algebra: In grades 5-8, the mathematics curriculum should include explorations of algebraic concepts of variable, expression, and equation.
MS (1993)	Mississippi defines skills for math education, with 20 to 30 skills per grade (K-8), and 20 to 50 skills per 9-12 course. For example, Grade 7: Solve word problems involving time using any one of the four operations. Note: Algebra content is not recommended at middle grade levels.
NH (1994)	New Hampshire defines 5 primary goals for grades K-12 with curriculum standards and proficiency standards given by grade level under each goal. For example, Functions, Relations, and Algebra: Grades 7-12: Building upon the K-6 experiences, analyze functions and relations to describe how a change in one quantity results in a change in another.
NJ (1993)	New Jersey defines 18 standards, including content, process and learning environment standards for K-12. For example, Standard 16: All students will develop their understanding of algebraic concepts and processes through experiences which enable them to describe, represent, and analyze relationships among variable quantities and to apply algebraic methods to solve meaningful problems.
SC (1993)	South Carolina defines content standards by grade block for each of the 6 core strands. Strands include 30 for grades K-3, 44 for grades 3-6, 50 for grades 6-9, and 65 for 9-12, including 23 for advanced mathematics. For example, Strand: Numerical and Algebraic Concepts and Operations (grades 6-9): Students will participate in problem-solving activities through group and individual investigations so that they can: Understand the concepts of variables, expressions, equations, and inequalities and gain confidence in thinking and communicating algebraically.
UT (1993)	Utah defines content standards for each grade and course with an average of 13 standards in grades 1-7, and 7 to 16 standards for grades 8-12. For example, Level 7, Standard 5070-09: The students will explore algebra concepts in an informal way to build a foundation for the subsequent formal study of algebra. 5070-0902: Analyze tables and graphs to identify properties and relationships.

BEST COPY AVAILABLE

Comparability of State Standards Documents — Reading

	ADJUST READING STRATEGIES TO MAKE SENSE OF TEXT	MAKE CONNECTIONS AMONG PIECES OF LITERATURE AS WELL AS LITERATURE AND SOCIETY	USE REFERENCE TOOL SOURCES	READ AND COMPREHEND A VARIETY OF TEXTS	USE/APPLY MULTIPLE DATA SOURCES	MAKE CONNECTIONS BETWEEN TEXT INCIDENTS/ ITEMS AND PERSONAL EXPERIENCE	INTERPRET TEXTS AND SUPPORT INTERPRETATIONS	EVALUATE TEXTS	ORGANIZE/ APPLY INFORMATION GATHERED	GENERATE QUESTIONS (WHILE, OR AS A RESULT OF, READING)	SELECT AND READ TEXT FOR A VARIETY OF PURPOSES
CO	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
DE	✔	✔	✔	✔	✔	✔	✔	✔	✔		✔
ID	✔		✔					✔		✔	✔
KS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
KY			✔	✔	✔	✔		✔	✔	✔	✔
LA	✔		✔	✔	✔	✔		✔	✔		✔
MA	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
MI	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
MO		✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
MS		✔	✔	✔	✔	✔		✔	✔	✔	✔
ND	✔	✔	✔			✔				✔	
NH	✔		✔	✔		✔	✔		✔	✔	✔
NJ	4	4	4	4	4	4	4	4	4	4	4
OR	✔	✔	✔	✔	✔	✔	✔	✔	✔		
RI	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
SC	✔	✔	✔	✔	✔	✔		✔	✔		✔
TN	✔		✔		✔		✔	✔	✔		✔
TX	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
VA	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
VT	✔	✔		✔	✔	✔	✔	✔	✔	✔	✔
WA	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
WI			✔	✔	✔				✔		
WV							✔	✔			

Comparability of State Standards Documents— Reading (continued)

	DRAW CONCLUSIONS	USE PRIOR KNOWLEDGE TO COMPREHEND	ANALYZE TEXTS AND INFORMATION SOURCES	USE THE STRUCTURE OF TEXTS TO COMPREHEND	RESPOND APPRECIATIVELY TO LITERATURE	DETERMINE PURPOSES FOR READING	IDENTIFY STORY ELEMENTS	MAKE AND CONFIRM PREDICTIONS	SUMMARIZE TEXT	UNDERSTAND, APPLY AND USE SOUND/SYMBOL CORRESPONDENCE	IDENTIFY MAIN IDEA	PRODUCE MULTIPLE RESPONSES TO TEXT
CO	✔	✔	✔	✔	✔		✔		✔	✔	✔	✔
DE	✔	✔	✔		✔	✔	✔	✔	✔			✔
ID					✔		✔		✔	✔	✔	
KS	✔	✔	✔		✔	✔	✔	✔	✔		✔	✔
KY		✔	✔			✔						
LA	✔	✔	✔	✔		✔	✔	✔	✔	✔	✔	
MA		✔	✔	✔		✔	✔	✔		✔		✔
MI	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔		✔
MO		✔	✔	✔								
MS		✔	✔		✔							
ND		✔	✔	✔		✔	✔	✔	✔	✔	✔	✔
NH		✔		✔		✔	✔	✔	✔	✔	✔	
NJ	✔	✔	✔	✔	✔	✔	✔		✔	✔	✔	✔
OR		✔	✔	✔	✔				✔		✔	✔
RI	✔				✔							
SC		✔	✔			✔	✔			✔		
TN			✔	✔		✔	✔		✔	✔		
TX	✔	✔	✔	✔		✔	✔	✔	✔	✔	✔	✔
VA	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	
VT				✔	✔	✔		✔		✔		
WA	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	
WI										✔		
WV	✔	✔	✔			✔	✔	✔	✔	✔	✔	

Comparability of Standards — Writing

	USE STANDARD CONVENTIONS OF LANGUAGE	WRITE/ COMPOSE USING VARIED FORMS AND PURPOSES	EVALUATE WRITTEN COMPOSITION	ORGANIZE THOUGHTS FOR WRITTEN COMPOSITION	UNDERSTAND THE RECURSIVE NATURE OF THE WRITING PROCESS	WRITE FOR MULTIPLE AUDIENCES	MATCH FORMS OF WRITING WITH PURPOSES FOR WRITING	WRITE AS A MEANS OF EXPRESSING & UNDERSTANDING SELF	DEMONSTRATE AN INDIVIDUAL STYLE AND VOICE	ENGAGE IN EDITING (FOR FINAL PUBLICATION)	ENGAGE IN REVISION	IDENTIFY A TOPIC
CO	✔	✔	✔	✔		✔		✔		✔	✔	✔
DE	✔	✔		✔	✔	✔	✔	✔	✔			
ID	✔	✔	✔							✔		✔
KS	✔	✔	✔	✔	✔	✔	✔		✔	✔	✔	
KY	✔	✔				✔	✔					
LA	✔	✔	✔	✔		✔	✔					
MA	✔	✔	✔	✔		✔	✔	✔		✔	✔	✔
MI	✔	✔	✔		✔	✔	✔		✔	✔	✔	
MO	✔	✔	✔	✔							✔	
MS	✔	✔	✔	✔		✔				✔	✔	
ND	✔	✔	✔	✔		✔	✔	✔		✔	✔	✔
NH	✔	✔	✔	✔		✔	✔			✔	✔	✔
NJ	✔	✔	✔	✔	✔	✔	✔	✔		✔	✔	✔
OR	✔	✔	✔	✔		✔	✔					
RI	✔	✔	✔		✔	✔		✔				
SC	✔	✔	✔	✔		✔	✔	✔	✔	✔	✔	✔
TN	✔	✔	✔	✔		✔	✔			✔	✔	
TX	✔	✔	✔	✔	✔	✔	✔	✔		✔	✔	
VA	✔	✔	✔		✔				✔			
VT	✔	✔	✔	✔			✔		✔	✔	✔	✔
WA	✔	✔	✔	✔		✔		✔		✔	✔	✔
WI	✔			✔								
WV	✔	✔				✔				✔	✔	✔

Descriptive Characteristics — English Language Arts Standards

	NUMBER OF LANGUAGE ARTS STANDARDS (BROADLY DEFINED OUTCOMES OR GOALS)	GRADE LEVEL BENCHMARKS	INCLUSION OF CONCEPTUAL FRAMEWORK IN DOUCMENT	INTEGRATION OF LANGUAGE ARTS	INTEGRATION OF THE CONTENT AREAS	DEFINITION OF TERMS	ILLUSTRATIONS, VIGNETTES, EXAMPLES	PLAN FOR PROFESSIONAL DEVELOPMENT	DESCRIPTION OF DEVELOPMENT PROCESSES	DRAFT VERSUS FINAL DOCUMENT	CONSISTENCY WITHIN/ BETWEEN CATEGORIES	PRESENCE OF ADDITIONAL RESOURCES/ BIBLIOGRAPHY
CO	6	K-4, 5-8, 9-12	yes	medium	low	no	not included	not specified	yes	final	high	yes
DE	4	K-3, 4-5, 6-8, 9-10	yes	high	high	no	vignettes	not specified	yes	final	high	yes
ID		K, 1, 2, 3, 4, 5, 6	no	low	low	no	not included	not specified	yes	final	high	no
KS	5	elementary, middle, secondary	no	medium	low	no	vignettes	not specified	no	draft	high	yes
KY	6	elementary, middle school, and high school demonstrations	no	medium	high	yes	yes	yes (minimal)	yes	final	high	yes
LA	6	K-4, 5-8, 9-12	yes	medium	low	yes	not included	not specified	yes	draft	high	yes
MA	6	PK-4, 5-8, 9-10, 11-12	yes	high	high	no	yes	not specified	no	final	high	yes
MI	12	early elementary, later elementary, middle school, high school	yes	high	high	no	yes	not specified	yes	final	high	no
MO	4	—	yes	high	high	no	not included	not specified	yes	final	high	no
MS	7	K-3, 4-8, 9-12	yes	high	low	yes	yes	no	yes	final	high	no
ND	6	4, 8, 12	no	medium	uncertain	yes	suggested activities	not specified	yes	final	medium	yes

BEST COPY AVAILABLE

Descriptive Characteristics — English Language Arts Standards (continued)

	NUMBER OF LANGUAGE ARTS STANDARDS (BROADLY DEFINED OUTLINES OR GOALS)	GRADE LEVEL BENCHMARKS	INCLUSION OF CONCEPTUAL FRAMEWORK IN DOCUMENT	INTEGRATION OF LANGUAGE ARTS	INTEGRATION OF THE CONTENT AREAS	DEFINITION OF TERMS	ILLUSTRATIONS, VIGNETTES, EXAMPLES	PLAN FOR PROFESSIONAL DEVELOPMENT	DESCRIPTION OF DEVELOPMENT PROCESSES	DRAFT VERSUS FINAL DOCUMENT	CONSISTENCY WITHIN/ BETWEEN CATEGORIES	PRESENCE OF ADDITIONAL RESOURCES/ BIBLIOGRAPHY
NH	7	3, 6, 10	yes	low	low	no	no	not specified	yes	final	high	yes
NJ	5	4, 8, 12	yes	medium	high	yes	not included	not specified	no	final	high	no
OR	5	3, 5, 8, 10, 12 (in progress)	no	low	low	no	not included	not specified	no	draft	high	no
RI	9	developmental levels K-4	yes	high	high	yes	not included	yes	yes	draft	high	yes
SC	4	PK-3, 4-5 6-8, 9-12	yes	high	medium	yes	vignettes	yes	yes	final	medium	yes
TN	4	K-2, 3-5, 6-8 9, 10, 11, 12	yes	low	low	yes	not included	not specified	yes	final	high	yes
TX	99	K-3, 4-8, 9-12	no	medium	low	no	not included	not specified	no	final	low	no
VA	uncertain*	K-12	yes	low	low	no	not included	not specified	no	draft	low	no
VT	5	K-4, 5-8, 9-12	no	high	high	yes	not included	not specified	yes	final	high	yes
WA	12	4, 7, 10	yes	low	low	yes	not included	not specified	yes	final	medium	no
WI	11	not yet developed	no (references national standards)	medium	low	no	not included	not specified	no	draft	low	yes
WV	4	by grade level K-8, high school	yes	low	low	no	not included	not specified	yes	final	high	no

*4 major categories, but multiple, distinct descriptors under each category at each grade level.

BEST COPY AVAILABLE

Comparing the Terminology of English Language Arts Standards Documents

	TERMS FOR THE PHILOSOPHICAL/ ORGANIZATIONAL UNDERPINNINGS OF THE DOCUMENT	TERMS FOR BROADLY DEFINED "OUTCOME" OR GOAL	TERMS FOR EXPECTATIONS FOR STUDENT LEARNING	TERMS FOR LEVELS OR DEGREE OF PERFORMANCE STUDENTS WILL ACHIEVE	TERMS FOR THE DESCRIPTIONS OF THE STANDARDS IN ACTION
CO	rationale (additional, more general philosophical statements found in introduction)	model content standard	no term noted — text reads — "in order to meet this standard students will...." (bullets)	Grade level bands K-4, 5-8, 9-12	none listed
DE	guiding principles and philosophy statements	content standards	unlabeled: elaborates content standards. (See pp. 2, 18, 34, 44).	performance indicators	vignettes
ID	not included	no broad goals	target skills	factual, applied, analysis synthesis/evaluation	none listed
KS	curricular standards or communications	outcome	standard	benchmarks	vignettes
KY	philosophy and underlying assumptions	learning goals	academic expectations	elementary, middle school and high school demonstrators	sample teaching strategies
LA	information literacy model for lifelong learning	standards	benchmarks	grade level bands (K-4, 5-8, 9-12)	none listed
MA	no title given, philosophy before each content strand	content strands (media, language, literature, composition)	learning standard	standards listed in grade level bands PK-4, 5-8, 9-10, 11-12	examples of sample literature units
MI	not labeled — but found in italicized language following each standard	English language arts content Standards	description or elaborations of the content standards	benchmarks (early elementary, later elementary, middle school, high school)	vignettes of units of study accompany standard document (each activity correlated to a particular content standard
MO	the Show-Me standards	goals (knowledge)	listed under each goal are statements "the students will demonstrate...."	none listed	none listed
MS	philosophy statement/ preamble	K-12 Goals	scope and sequence of competencies	K-3 competencies 4-8 competencies 9-12 competencies	suggested teaching strategies

BEST COPY AVAILABLE

Comparing the Terminology of English Language Arts Standards Documents (continued)

	TERMS FOR THE PHILOSOPHICAL/ ORGANIZATIONAL UNDERPINNINGS OF THE DOCUMENT	TERMS FOR BROADLY DEFINED "OUTCOME" OR GOAL	TERMS FOR EXPECTATIONS FOR STUDENT LEARNING	TERMS FOR LEVELS OR DEGREE OF PERFORMANCE STUDENTS WILL ACHIEVE	TERMS FOR THE DESCRIPTIONS OF THE STANDARDS IN ACTION
ND	not present	content standards	none listed	benchmarks	activities
NH	purpose statements	curriculum standards	curriculum and proficiency standards listed in terms "students will be able to...." (end of grade *x*)	benchmarks listed "by the end of grades 3, 6, 10."	none listed
NJ	descriptive statements and introductions	core curriculum content	cumulative progress indicators	"by the end of grades 4, 8, and 12"	none listed
OR	common curriculum goals	content standard (found in boldface type in document)	content standards (listed as bullets)	benchmarks	none listed
RI	critical beliefs about English language arts	content standards	descriptors	four developmental levels of indicators (level 1, 2, 3, and 4)	none listed
SC	introduction to the framework	strand	standard	descriptors (listed by grade level bands PK-3, 4-5, 9-12)	Vignettes
TN	goal statement	content standard	learning expectations	K-2, 3-5, 6-8, 9, 10, 11, 12	none listed
TX	basic understanding	essential knowledge and skill	performance descriptor	grade level bands (K-3, 4-8, 9-12)	classroom stories (not yet developed)
VA	goal of English language arts (p. 1)	not stated as "goals," standards organized into four categories: oral language, reading/literature, writing and research.	English standards of learning (numbered and found in boldface type)	English standards of learning (listed as bullets)	overview (at each grade level)
VT	not present	standards organized into four categories: science, math and technology; humanities; history and social sciences; communication, reasoning and problem solving, personal development, civic/social responsibility	evidence	grade level cluster	none listed

BEST COPY AVAILABLE

Comparing the Terminology of English Language Arts Standards Documents (continued)

	TERMS FOR THE PHILOSOPHICAL/ ORGANIZATIONAL UNDERPINNINGS OF THE DOCUMENT	TERMS FOR BROADLY DEFINED "OUTCOME" OR GOAL	TERMS FOR EXPECTATIONS FOR STUDENT LEARNING	TERMS FOR LEVELS OR DEGREE OF PERFORMANCE STUDENTS WILL ACHIEVE	TERMS FOR THE DESCRIPTIONS OF THE STANDARDS IN ACTION
WA	introduction	essential academic learning requirement	components	benchmarks (4, 7, 10)	none listed
WI	English language arts within a framework connecting arts, social studies, foreign language, and ELA	program standards	content standards	performance standards	none listed
WV	assumptions (pp. 1-2)	English language arts program of study K-12 (sections 4-7)	instructional goals	objectives (levels broadly defined: early childhood, middle childhood, adolescent education)	none listed

BEST COPY AVAILABLE

English Language Arts Content Standards

	SPECIFICITY/MEASURABILITY		
	HIGH	MEDIUM	LOW
CO	✔		
DE	✔		
ID		✔	
KS			✔
KY	✔		
LA		✔	
MA	✔		
MI		✔°	
MO			✔
MS		✔	
MT		✔	
ND		✔	
NH	✔		
NJ		✔	
OR		✔	
RI	✔		
SC		✔°	
TN	✔		
TX	✔		
VA	✔		
VT	✔		
WA	✔		
WI			✔
WV	✔		

°Statements varied in terms of specificity.

BEST COPY AVAILABLE

HIGH — Provides explicit explanation as to the nature of the expected performance and the guidelines for establishing the context in which students can demonstrate the standard. Examples:

- ☑ "Critically analyze and evaluate information and messages presented through print, speech, and mass media by recognizing a variety of persuasive and propaganda techniques and how they are used in a variety of forms (e.g., advertising, campaigns, and news formats)."
- ☑ "Use correct spelling for frequently used words and phonetically regular words in final copies."

MEDIUM — Provides *some* definition and direction for determining how students can demonstrate the standard. Examples:

- ☑ "Identify a topic for a composition and determine its development."
- ☑ "Make and confirm simple predictions to increase understanding."

LOW — Fails to convey enough detail or definition to produce an interpretable "view" of what a learner would actually be doing while demonstrating this standard. Examples:

- ☑ "The learners will prepare writing and speaking for the world outside their classroom."
- ☑ "Demonstrate an understanding that language is symbolic, arbitrary, and systematic."

BEST COPY AVAILABLE

The Analysis Process for Reading and Writing Standards

- ✔ The purpose of this review was to provide readers with an overview and categorical description of as many available state standards documents as possible.
- ✔ Reviewers did not evaluate the quality of each document. Reviewers simply indicated whether certain information was contained in the document.
- ✔ The absence of information /checks in a column does not necessarily indicate an incomplete or inferior document. For example, some states chose to be more global in their language, clearly leaving many decisions at the district level. Other states provided a highly detailed scope and sequence that resulted in checks in all categories.
- ✔ Some states indicated that their documents were the result of broad-based consensus of language arts educators across the state; other documents seemed to be the work of state legislators in combination with state department staff. Questions remain across all documents about the impact of a single individual or powerful special interest groups. Information on the development of the documents was absent or incomplete in some documents.
- ✔ Some documents were final and approved by the state's board of education; others were in draft form. Some documents may have been incomplete. For example, the information from one state contained only the elementary curriculum in English language arts.
- ✔ Categories emerged from careful examination of a wide range of state documents. The first attempt at categorization resulted in broader language than the language now seen in this overview of state documents. For example, such categories as "inquiry" and "connections" were considered. When broader categories were used, however, all documents tended to look the same. The more specific categories, while limiting in some ways, tended to reveal some differences across states.
- ✔ Categories were adhered to carefully. For example, many states listed the "steps" found in the writing process. In fact, most states are given credit on the chart for attention to revision and editing. However, fewer states are given credit for attending to the "recursive nature of the writing process" because they did not mention the flexible and interactive nature of the steps found within the

process. While this information may be in supplemental state standards documents, it was not available in many of the standards documents examined.

- ✔ Rather than match exact wording of categories to words found in the state documents, reviewers carefully considered intent as well as wording.

Bibliography

Colorado State Board of Education. (July 13, 1995). *Colorado Model Content Standards for Reading and Writing*. Information accessed October 14, 1997. Available http://www.cde.state.co.us/read.htm. Denver, CO.

Commonwealth of Virginia Department of Education. (June 1995). *English Standards of Learning*. Richmond, VA.

Delaware Department of Public Instruction. (June 1995). *New Directions: Delaware English Language Arts Curriculum Framework Content Standards Volume 1*. (Document Control Number 95-01/95/05/05). Information accessed October 14, 1997. Available http://www.dpi.state.de.us/dpi/standard/engla/eindex.html. Dover, DE.

Idaho State Department of Education. (1990, Reprinted April 1995). *Integrated Language Arts Course of Study K-8*. Boise, ID.

Idaho State Department of Education. (April 1995). *Integrated Language Arts Model K-12*. Boise, ID.

Kansas State Board of Education. (October 1996). *Kansas Curricular Standards for Communication Arts*. Information accessed September 27, 1997. Available http://www.ksbe.state.ks.us/outcomes/Commtble.html. Topeka, KA.

Kentucky Department of Education. (September 26, 1997). *Kentucky Learning Goals and Academic Expectations*. Information accessed September 26, 1997. Available http://www.kde.state.ky.us. Frankfort, KY.

Louisiana Department of Education. (1997 Draft). *Schools That Work: Setting Higher Standards for Our Students. Louisiana English Language Arts Content Standards*. Baton Rouge, LA.

Massachusetts Department of Education. (February 1997). *Content Strands and Learning Standards in the English Language Arts Curriculum Framework*. Information accessed September 26, 1997. Available http://www.doe.mass.edu/doedocs/framework/englishCS.html. Malden, MA.

Michigan Department of Education. (April 1995, draft). *English Language Arts: Content Standards and Working Draft Benchmarks*. Information accessed October 14, 1997. Available http://cdp.mde.state.mi.us/ContentStandards/EnglishLanguageArts/. Lansing, MI.

Mississippi Department of Education. (1996). *Mississippi Language Arts Framework*. Jackson, MS.

Missouri Department of Elementary and Secondary Education. (February 1996).

The Show-Me Standards: Knowledge + Performance = Academic Success. Jefferson City, MO.

New Hampshire Department of Education. (September 27, 1997). *New Hampshire K-12 English Language Arts Curriculum Framework*. Information accessed September 27, 1997. Available http://www.state.nh.us/doe/engfull.htm#reading. Concord, NH.

New Jersey Department of Education. (February 1996). *Core Curriculum Standards in Language Arts and Literacy*. Information accessed September 27,1997. Available http://www.injersey.com/Education/NJDOE/08langstan3_1.html. Trenton, NJ.

North Dakota Department of Public Instruction. (1994). *North Dakota English Language Arts Curriculum Frameworks: Standards and Benchmarks*. Bismarck, ND.

Oregon Department of Education. (1997). *Common Curriculum Goals: Academic Content Standards, K-12 Benchmarks & Essential Learning Skills*. Information accessed September 27, 1997. Available http://www.ode.state.or.us/inst/cfsi/ccg.htm. Salem, OR.

Rhode Island Department of Elementary and Secondary Education. (October 1995, draft). *Literacy for ALL Students: The Rhode Island English Language Arts Framework*. Information accessed September 27, 1997. Available http://instruct.ride.ri.net/ride_home_page.html. Providence, RI.

Tennessee State Department of Education. (December 5, 1996). *English Language Arts Curriculum Framework K - 12: Writing, Reading, Viewing and Representing, Speaking and Listening*. Nashville, TN.

Texas Education Agency. (July 1997). *English Language Arts and Reading: Texas Essential Knowledge and Skills*. Information accessed September 27,1997. Available http://www.tea.state.tx.us./teks/index2.html. Austin, TX.

Vermont Department of Education. (July 1994, draft). *First Draft Content Standards: A first Step Toward Vermont's Common Core Framework for Curriculum and Assessment*. Information accessed September 27, 1997. Available http://www.state.vt.us/educ/stand/page3.htm. Montpelier, VT.

Washington State Commission on Student Learning. (1997). *Essential Academic Learning Requirements: Reading, Writing, Communication*. Information accessed September 27, 1997. Available http://www.ospi.wednet.edu. Olympia, WA.

Wisconsin Department of Public Instruction. (October 1997 draft; final copy expected December 1997). *Wisconsin Academic Standards for English Language Arts*. Information accessed September 27, 1997. Available http://www.state.wi.us/agencies/api/standards/. Madison, WI.

West Virginia Department of Education. (1997). *West Virginia Instructional Goals and Objectives*. Information accessed September 27, 1997. Available http://access.k12.wv.us. Charleston, WV.

BEST COPY AVAILABLE

Websites

The Council of Chief State School Officers provides direct access to all available state education agency websites at: http://www.ccsso.org/seamenu.html.

COLORADO STATE BOARD OF EDUCATION
http://www.cde.state.co.us/read.htm

COMMONWEALTH OF VIRGINIA DEPARTMENT OF EDUCATION
http://pen.k12.va.us/anthology/VDOE

DELAWARE DEPARTMENT OF PUBLIC INSTRUCTION
http://www.dpi.state.de.us/dpi/standard/engla/eindex.html

IDAHO STATE DEPARTMENT OF EDUCATION
http://www.sde.state.id.us/Dept

KANSAS STATE BOARD OF EDUCATION
http://www.ksbe.state.ks.us/outcomes/Commtble.html

KENTUCKY DEPARTMENT OF EDUCATION
http://www.kde.state.ky.us

LOUISIANA DEPARTMENT OF EDUCATION
http://www.doe.state.la.us

MASSACHUSETTS DEPARTMENT OF EDUCATION
http://www.doe.mass.edu/doedocs/framework/englishCS.html

MICHIGAN DEPARTMENT OF EDUCATION
http://cdp.mde.state.mi.us/ContentStandards/EnglishLanguageArts/

MISSISSIPPI DEPARTMENT OF EDUCATION
http://mdek12.state.ms.us

MISSOURI DEPARTMENT OF ELEMENTARY AND SECONDARY EDUCATION
http://www.cde.state.co.us/read.htm

NEW HAMPSHIRE DEPARTMENT OF EDUCATION
http://services.dese.state.mo.us

NEW JERSEY DEPARTMENT OF EDUCATION
http://www.injersey.com/Education/NJDOE/08langstan3_1.html

NORTH DAKOTA DEPARTMENT OF PUBLIC INSTRUCTION
http://www.dpi.state.nd.us

OREGON DEPARTMENT OF EDUCATION
http://www.ode.state.or.us/inst/cfsi/ccg.htm

RHODE ISLAND DEPARTMENT OF ELEMENTARY AND SECONDARY EDUCATION
http://www.instruct.ride.ri.net/ride_home_page.html

BEST COPY AVAILABLE

TENNESSEE STATE DEPARTMENT OF EDUCATION
http://www.state.tn.us/eduction

TEXAS EDUCATION AGENCY
http://www.tea.state.tx.us./teks/index2.html

VERMONT DEPARTMENT OF EDUCATION
http://www.state.vt.us/educ/stand/page3.htm

WASHINGTON STATE COMMISSION ON STUDENT LEARNING
http://www.ospi.wednet.edu

WISCONSIN DEPARTMENT OF PUBLIC INSTRUCTION
http://www.state.wi.us/agencies/api/standards/

WEST VIRGINIA DEPARTMENT OF EDUCATION
http://access.k12.wv.us

BEST COPY AVAILABLE

Descriptive Characteristics — Social Studies Standards — Part I

	NUMBER OF SOCIAL STUDIES STANDARDS	GRADE LEVEL BENCHMARKS	CONCEPTUAL FRAMEWORK	INTEGRATION OF THE CONTENT AREAS	DEFINITION OF TERMS	EXAMPLES, ILLUSTRATIONS VIGNETTES
AK	16	ages 7, 10, 14, 18	yes	high	no	activities
AL	by grade levels	yes	yes	low	yes	yes
AR	6	4, 8, 12	yes	high	no	—
AZ			*under development*			
CA	by grade levels	K-12	yes	high	yes	yes
CO	20	4, 8, 12	yes	low	yes	yes
CT	15	4, 8, 12	yes	high	yes	yes
DC	7 (history)	3, 5, 7, 11	yes	low	no	no
DE	16	3, 5, 8, 12	yes	low	no	yes
FL	11	2, 5, 8, 12	yes	high	no	no
GA			*not available*			
HI	by grade levels	3, 6, 8, 12	yes	high	yes	yes
IA				*not developing standards*		
ID	by grade levels	4-8-exit	yes	high	no	yes
IL	20	ee/le-m-eh/lhs	yes	high	no	no
IN	by grade levels	K-12	yes	high	yes	yes
KS	6	5, 8, 12	yes	high	no	yes
KY	5	E-M-H	yes	high	no	no
LA	4	4, 8, 12	yes	low	no	no
MA	6	(4, 8, 9, 12)	yes	high	no	yes
MD	7	(3, 5, 8, HS)	yes	high	no	no
ME	18	(2-4-8-secondary)	yes	low	no	no
MI	25	(ee-le-ms-hs)	yes	high	yes	yes
MN	5	no	yes	low	no	yes

BEST COPY AVAILABLE

Descriptive Characteristics — Social Studies Standards — Part I (continued)

	NUMBER OF SOCIAL STUDIES STANDARDS	GRADE LEVEL BENCHMARKS	CONCEPTUAL FRAMEWORK	INTEGRATION OF THE CONTENT AREAS	DEFINITION OF TERMS	EXAMPLES, ILLUSTRATIONS VIGNETTES
MO	7	(4, 8, 12)	yes	high	yes	yes
MS	grade level courses		yes	low	no	no
MT			*under development*			
NC			*not available*			
ND	14	4, 8, 12	yes	low	no	no
NE			*not available*			
NH	18	(6, 10)	yes	high	no	no
NJ	9	(4, 8, 12)	yes	low	no	no
NM			*under development*			
NV			*not available*			
NY	6	(Enter, commencement)	yes	high	no	no
OH	3	K-12	yes	high	yes	yes
OK	by grade levels		yes	no	no	no
OR	3	5, 8, 10	yes	low	no	no
PA			*under development*			
RI			*not developing standards*			
SC			*not available*			
SD	18	2, 4, 8, 12	yes	low	yes	no
TN	18	2, 5, 8, 12	yes	low	no	no
TX	by grade levels		no	low	no	yes
UT	by grade levels	K-12	yes	low		
VA	by grade levels	5, 8, 12	yes	low	no	no
VT	25	4, 8, 12	yes	high	no	no

BEST COPY AVAILABLE

Descriptive Characteristics — Social Studies Standards — Part I (continued)

	NUMBER OF SOCIAL STUDIES STANDARDS	GRADE LEVEL BENCHMARKS	CONCEPTUAL FRAMEWORK	INTEGRATION OF THE CONTENT AREAS	DEFINITION OF TERMS	EXAMPLES, ILLUSTRATIONS VIGNETTES
WA	12	flexible	yes	high	no	no
WI	10	4, 8, 12	yes	yes	no	no
WV	grade levels		yes	low	no	no
WY	*standards being developed by local school systems*					

Descriptive Characteristics — Social Studies Standards — Part II

	PLAN FOR PROFESSIONAL DEVELOPMENT	DESCRIPTION OF DEVELOPMENT PROCESS	DYNAMIC VS FINAL DOCUMENT	CONSISTENCY WITHIN/BETWEEN CATEGORIES	PRESENCE OF ADDITIONAL RESOURCES/ BIBLIOGRAPHY
AK	guidelines for LEAs	yes	dynamic	high	yes
AL	no	no	final	high	yes
AR			dynamic	high	
AZ	*under development*				
CA	no	no	final	high	yes
CO	no	no	dynamic	high	index: terms and topics
CT	yes	yes	dynamic	high	yes
DC	no	no	dynamic	high	no
DE	no	no	final	high	no
FL	no	no	final	high	no
GA	*not available*				
HI	no	yes	final	high	yes
IA	*not developing standards*				
ID	no	no	final	high	yes
IL	no	no	dynamic	high	yes
IN	no	no	final	high	yes

Descriptive Characteristics — Social Studies Standards — Part II (continued)

	PLAN FOR PROFESSIONAL DEVELOPMENT	DESCRIPTION OF DEVELOPMENT PROCESS	DYNAMIC VS FINAL DOCUMENT	CONSISTENCY WITHIN/BETWEEN CATEGORIES	PRESENCE OF ADDITIONAL RESOURCES/ BIBLIOGRAPHY
KS	no	no	final	high	yes
KY	no	no	dynamic	high	no
LA	no	no	dynamic	high	no
MA	no	no	dynamic	high	no
MD	no		dynamic	high	yes
ME	no	no	dynamic	high	no
MI	no	no	dynamic	high	yes
MN	no	no	dynamic	high	yes
MO	no	no	dynamic	high	no
MS	no	no	final	high	no
MT			*under development*		
NC			*not available*		
ND	no	no	final	high	no
NE			*not available*		
NH	no	no	dynamic	high	yes
NJ	no	no	dynamic	high	no
NM			*under development*		
NV			*not available*		
NY	no	no	dynamic	high	yes
OH	no	no	final	high	yes
OK	no	no	final	high	no
OR	no	no	dynamic	high	no
PA			*under development*		
RI			*not developing standards*		

BEST COPY AVAILABLE

Descriptive Characteristics — Social Studies Standards — Part II (continued)

	PLAN FOR PROFESSIONAL DEVELOPMENT	DESCRIPTION OF DEVELOPMENT PROCESS	DYNAMIC VS FINAL DOCUMENT	CONSISTENCY WITHIN/BETWEEN CATEGORIES	PRESENCE OF ADDITIONAL RESOURCES/ BIBLIOGRAPHY
SC			*not available*		
SD	no	no	final	high	yes
TN	no	no	dynamic	high	no
TX	no	no	dynamic	high	no
UT	no	yes	dynamic	high	yes
VA	no	no	dynamic	yes	no
VT	no	yes	dynamic	high	yes
WA	no	no	final	yes	no
WI	no	no	final	high	no
WV	no	no	final	high	no
WY			*standards being developed by local school systems*		

BEST COPY AVAILABLE

Comparing the Terminology of Social Studies Standards Documents

	TERMS FOR THE PHILOSOPHICAL/ ORGANIZATIONAL UNDERPINNINGS OF THE DOCUMENT	TERMS FOR BROADLY DEFINED OUTCOME OR GOAL	TERMS FOR EXPECTATIONS FOR STUDENT LEARNING	TERMS FOR LEVELS OR DEGREE OF PERFORMANCE STUDENTS WILL ACHIEVE	TERMS FOR THE DESCRIPTIONS OF THE STANDARDS IN ACTION
AK	integrating the social science disciplines	standards	a student who meets the standard should:	level(s)	activities
AL	introduction	topics	student outcomes	outcomes	outcomes
AR	strands	content standard	student learning expectations		
AZ			under development		
CA	introduction	goal	strands		
CO	purpose	content standard	know and be able to do		
CT	goals	goals	standards	performance standards	sample activities on illustrative tasks and selected samples of student work
DC	content standards (history only)	student performance indicators	what students should know and be able to do		
DE	guiding principles	content standards			
FL	strand		standard	benchmark	the student knows or understands
GA			*not available*		
HI	rationale	content standards	performance standards		
IA			*not developing standards*		
ID	goals	themes	performance objectives		
IL	goals	standards	standards		
IN	rationale	content strands	proficiency statements and indicators		
KS	purposes and principles	program outcomes	program outcomes	standards	benchmarks
KY		content guidelines	core content		
LA	social studies strands		content standards		
MA	guiding principles		social studies content	learning strands	questions
MD	outcomes		expectations of learning	indicators	

Comparing the Terminology of Social Studies Standards Documents (continued)

	TERMS FOR THE PHILOSOPHICAL/ ORGANIZATIONAL UNDERPINNINGS OF THE DOCUMENT	TERMS FOR BROADLY DEFINED OUTCOME OR GOAL	TERMS FOR EXPECTATIONS FOR STUDENT LEARNING	TERMS FOR LEVELS OR DEGREE OF PERFORMANCE STUDENTS WILL ACHIEVE	TERMS FOR THE DESCRIPTIONS OF THE STANDARDS IN ACTION
ME	guiding principles	learning results	standards		
MI	framework	strand	standards		benchmarks
MN	high standards	standards	content standards	what students should know	
MO	strands	content requirements	what all students should know	what students should know and be able to do	
MS	goal	strand/strand goal	core objective		
MT			*under development*		
NC			*not available*		
ND	mission statement	content	outcomes statement	benchmarks/performance standards	
NE			*not available*		
NH	purpose	curriculum standard	proficiency standard		
NJ	introduction	themes	standards	cumulative progress indicators	
NM			*not available*		
NV			*not available*		
NY	rationale	trends and issues	standards	performance indicators	
OH	mission	goal	objective		the learner will
OK	overview	academic	student skills		the student will
OR	common curriculum goals	content standards		benchmarks	
PA			*under development*		
RI			*not developing standards*		
SC			*not available*		
SD	vision statement		standards rationale	benchmarks	
TN	rationale	standard	standard		
TX		basic understanding	knowledge and skills		performance descriptions
UT	core standards	standards	objectives		

Comparing the Terminology of Social Studies Standards Documents (continued)

	TERMS FOR THE PHILOSOPHICAL/ ORGANIZATIONAL UNDERPINNINGS OF THE DOCUMENT	TERMS FOR BROADLY DEFINED OUTCOME OR GOAL	TERMS FOR EXPECTATIONS FOR STUDENT LEARNING	TERMS FOR LEVELS OR DEGREE OF PERFORMANCE STUDENTS WILL ACHIEVE	TERMS FOR THE DESCRIPTIONS OF THE STANDARDS IN ACTION
VA	goals		standards		
VT	vital results	critical evaluation	standards		
WA	goal	essential learning component		benchmarks	content examples
WI	rationale		performance standards		
WV	core and required programs of study		social studies instructional goals		
WY			*not developing standards*		

BEST COPY AVAILABLE

Social Studies Performance Standards

SEA	SPECIFICITY / MEASURABILITY		
	HIGH	MEDIUM	LOW
AK		✔	
AL	✔		
AR		✔	
AZ		*under development*	
CA	✔		
CO	✔		
CT	✔		
DC	✔		
DE	✔		
FL	✔		
GA		*not available*	
HI	✔		
IA		*not developing standards*	
ID		✔	
IL	✔		
IN	✔		
KS		✔	
KY		✔	
LA	✔		
MA		✔	
MD		✔	
ME	✔		
MI		✔	
MN		✔	
MO		✔	
MS	✔		
MT		*under development*	

BEST COPY AVAILABLE

Social Studies Performance Standards (continued)

	SPECIFICITY / MEASURABILITY		
SEA	HIGH	MEDIUM	LOW
NC		*not available*	
ND		✔	
NE			
NH	✔		
NJ			
NM		*under development*	
NV		*not available*	
NY		✔	
OH		✔	
OK		✔	
OR		✔	
PA		*under development*	
RI		*not developing standards*	
SC		*not available*	
SD		✔	
TN	✔		
TX	✔		
UT		✔	
VA	✔		
VT		✔	
WA		✔	
WI	✔		
WV		✔	
WY		*standards developed by local school systems*	

BEST COPY AVAILABLE

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

INTRODUCTION

4

IMPLEMENTATION MODELS AND PROTOTYPES

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

POLICY

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

CRITERIA FOR STANDARDS

The criteria proposed here were derived by a group representing several education organizations, as well as representatives from state education agencies and large urban districts. The group began with an analysis of several sets of proposed criteria that included the following:

- ✔ "Promises to Keep: Creating High Standards for American Students," Report to the National Education Goals Panel
- ✔ "Criteria for High Quality Standards," American Federation of Teachers
- ✔ "Developing Standards for Standards: How Can the Standards Certification Process Best Serve the Goal of Improving Teaching and Learning," The Pew Forum on Education Reform
- ✔ "The Challenge of Change: Standards to Make Education Work for All Our Children," National Alliance of Business
- ✔ "Recommended Criteria and Processes for Responding to State and Voluntary National Standards for Education," Council of Chief State School Officers
- ✔ "Report of the Higher Education Advisory Group of the National Education Goals Panel for the National Education Standards and Improvement Council"

After identifying common criteria across the documents, the group narrowed the set of criteria and discussed the appropriateness of each criterion at length. They also asked representatives of states and districts to review the criteria and meet with them to discuss issues and recommendations.

The resulting criteria represent a compromise among the various perspectives represented in the meetings.

These criteria are the result of many hours of working, discussing, drafting, responding, and redrafting on the part of representatives of several organizations committed to the improvement of teaching and learning through standards-focused education systems. The final criteria evolved through consensus and do not necessarily represent those most valued by each of these organizations. However, it is important to recognize and thank the organizations for their thoughtful contributions and diligent efforts. The organizations involved in the preparation of these criteria include the following:

BEST COPY AVAILABLE

- ✔ Alliance for Curriculum Reform
- ✔ American Federation of Teachers
- ✔ Carnegie Corporation of America
- ✔ Council for Basic Education
- ✔ Council of Chief State School Officers
- ✔ Council of Great City Schools
- ✔ National Alliance of Business
- ✔ National Education Goals Panel
- ✔ National Endowment for the Arts
- ✔ U. S. Department of Education

The criteria are meant to serve as a focus for formative evaluation of state content standards. It is difficult to judge the quality of content standards without considering the many aspects of their implementation from assessments to policies, the contexts in which they were developed, and the purposes that they are to serve. All of these factors vary across states. However, it is hoped that these criteria will assist in thoughtful considerations of what has been accomplished, how effective that work has been, and what remains to be done. The discussion questions that follow each criterion are suggested as a means of using the criteria for such a formative evaluation process.

BEST COPY AVAILABLE

CRITERIA FOR HIGH QUALITY STANDARDS: A CONSENSUS AND CONSOLIDATION

Ever since the education reform movement was launched with the publication of *A Nation at Risk* in 1983, raising the quality of America's schools and boosting the academic performance of America's students have been top priorities for parents, educators, business people, and policy makers.

By the late 1980s, a broad and strong consensus emerged in support of setting clear, rigorous standards for what students should know and be able to do, and holding schools and students accountable for meeting those standards. In 1989, the President along with all 50 of the nation's governors, established a set of six education goals to be achieved in the United States by the year 2000. In the wake of this summit, educators began to develop voluntary national standards in the various content areas, and many states and districts launched their own standards-setting efforts. The signing of the Goals 2000 Act in 1994 provided valuable resources to states in their own standards-setting efforts. In March 1996, the education summit of governors and leading chief executive officers (CEOs) indicated a continuing, even strengthened, commitment to systemic, standards-focused education improvement.

At last, it seemed that the U.S. educational system would be based on a set of clearly defined expectations for students. Through setting standards at the state or local levels, parents, educators, business people, advocates of higher education, and policy makers would determine what was essential for students to know and be able to do — to ensure the economic prosperity of the nation as well as to maximize opportunities and quality of life for everyone.

All children, no matter where they lived in their state or their community, would face a challenging curriculum — as challenging as that faced by students in any other industrialized country. With clear standards, students would receive seamless instruction as they moved through increasingly advanced levels of education — or even as they transferred from school to school. Teachers would know the status of students' knowledge and proficiency in the beginning of each school year. Common standards would provide guidance to publishers of textbooks and would help to focus teacher training and other instructional development activities. With content standards in place, states and local districts could construct other components of an accountability system, most notably assessments — rigorous assessments of what students learn in the classroom and can apply beyond the classroom, rather than multiple choice instruments that test only a narrow range of skills.

Perhaps most important, making the goals and the end results of the educational process much more explicit for everyone would be a way to strengthen confidence in and support for the educational system. Students would know what they are expected to learn, and teachers would have more direction for what they should teach. Parents would have a much better sense of what schools are teaching their children and be able to judge whether their children have learned what they should. If students were meeting standards, the public would be more likely to lend support — both morally and financially — to their schools. If they are not, the public could legitimately demand changes in instructional practices and the use of additional resources. With clear goals and the means to measure progress toward them, states and districts could eliminate burdensome rules and regulations that inhibit innovation at the school level. Colleges and employers would have a clearer idea of the skills and knowledge that students should possess as they emerge from the system and, therefore, would be less likely to rely on standardized tests or (in the case of employers) reject new graduates altogether in favor of more experienced workers.

Despite these benefits and its broad acceptance, the standards movement has perhaps lost some of its early momentum in recent years. While many states and districts continue their standards-setting efforts, some states and districts have expressed a strong need for additional guidance and technical assistance before starting to develop or set standards themselves. In many states, educators, policy makers, and the public are grappling with the question of local control and the extent to which content decisions should be made at the state level. The lingering effects of the "outcomes-based education" movement, which fostered the public's fears that schools had abandoned the academic basics, have caused some to reject the whole notion of standards.

The Standards Criteria Consolidation Committee

It is in this context that the Standards Criteria Consolidation Committee was formed. The committee grew out of a larger group of education organizations, referred to as the Systemic Reform Strategic Planning Group, which was convened by the Council of Chief State School Officers. The group meets regularly in Washington, D.C., to exchange information and ideas. A number of the organizations in this larger group recognized that they could strengthen and advance the standards movement if they developed a common set of principles or "criteria" that state and local officials, educators, parents, and members of the

public could use to develope and evaluate their own standards. These criteria, it was hoped, would describe the characteristics that standards should possess in order to yield all the benefits described earlier and to allay some of the general concerns about standards. While sets of criteria had been developed by several groups, including the American Federation of Teachers, the Council of Chief State School Officers, the National Alliance of Business, the Pew Forum, and the National Education Goals Panel, each set had a somewhat different focus because of the varying agendas and different points in time for development.

As a result, representatives from various organizations began meeting regularly to develop a consolidated set of criteria. The Standards Criteria Consolidation Committee began by identifying criteria addressed repeatedly across the existing sets. They also developed a list of issues related to the evaluation of standards. After a number of meetings and several hours of spirited discussion, the group reached a consensus on a number of those issues. Reaching a consensus on other issues proved more complex, however. Drafts were written and rewritten. The committee eventually decided to draft a full document that would summarize the group's progress to date and could be circulated for discussion, comment, and trial in states and districts.

This final document is the result of consensus as well as consolidation and, thus, represents the narrowing of focus that comes with such efforts. Several organizations involved in the development of this document believe other criteria are important. Therefore, copies of their criteria are included as resources in this document.

The Content and Scope of this Document

This document contains six criteria for high quality standards. Some of the criteria are briefly stated and are accompanied by at most a brief explanation. Others — the more complex issues that were the focus of most of the Standards Criteria Consolidation Committee's discussions — are accompanied by much lengthier explanations. While these criteria do not represent an absolutely exclusive list, they are all necessary prerequisites for standards to be effective. They can serve as tools to guide states and districts through the standards-setting process toward systemic reform. States and districts can use these criteria to see what additional steps they need to take or whether they can improve what they have done so far. They may also use these criteria to decide whether to use other jurisdictions' standards as models or to draw upon documents produced by various national content groups.

This report contains criteria for high quality *content* standards. Promulgated at the state or local level, content standards are statements that describe the knowledge and skills students are expected to learn in school. Here the term *content standards* refers to any articulation of content at the state or district level.

Content standards are an essential component of systemic reform, but they make up only one step. They are the foundation upon which performance standards, assessments, and accountability measures are built. The committee focused primarily on content standards partly because the members wanted to limit the number of complex issues that they addressed and partly because many states and districts have completed drafts of content standards, but have not yet developed all the other components of systemic reform. However, other aspects of education improvement arise in various places in the document, when it is necessary to describe the characteristics that content standards must have to make those other components work effectively. In reality, content standards are not in a final, acceptable form until they have been actually implemented and aligned with the other components of systemic improvement.

Additional Qualities That Good Content Standards Possess

The criteria discussed in this document depend on certain widely held assumptions about standards-based reform. These assumptions affect both the actual content of standards and the way in which standards are developed and implemented in schools and communities. These basic conditions for reform should be made plain from the outset and kept in mind where they pertain to the following criteria. Good standards are linked to a careful, effective process of development and implementation. Consequently, it is difficult to judge the full quality of standards without considering the purposes for which they are developed, the contexts in which they are generated and implemented, and the resulting alignment of the various aspects of the education system. Because these considerations vary widely from state to state and district to district, it is important to contextualize the application of these criteria.

A Final Word on Background

This document is still a work in progress; it does not represent the final word of the Standards Criteria Consolidation Committee or the official position of any of the organizations that participated in this committee. In

some places, differences of opinion in the group have not yet been resolved, and arguments on different sides of the issue have simply been presented for the users of the document to consider for themselves.

CRITERION 1

Standards should expect and support all students achieving to high levels.

The desire for rigorous academic standards at the state and local level is based in part on the desire to ensure that all children — regardless of their environment — have a high quality curriculum and are challenged and supported to achieve to high levels. In the absence of rigorous standards for all students, some youngsters never have the opportunity to do more than the minimum. While this can be a problem in all schools and for students of all ethnicities and backgrounds, it disproportionately affects low income and minority students. All students should learn content and skills that are rich and challenging, and schools should be given the resources they need to have students achieve that learning. Content standards should promote that achievement.

Content standards must set forth a common body of knowledge that all students are expected to learn. This will help ensure that *all* students are exposed to a challenging core curriculum, no matter where they live or what school they attend. In addition, standards documents should take into account and build on differences in developmental levels, while recognizing that those levels are starting points and not limitations.

No matter how high content standards are set, some students will be capable of going beyond the common standards. This concern — that the standards may be too low to motivate higher achieving students — is being voiced loudly in some communities and in some cases has resulted in opposition to standards-based reform in general. The challenge before states and districts is to develop a system of standards that does not set any ceilings. Standards should motivate and challenge all students to high, substantive achievement.

In order to alleviate these concerns, content standards should eventually be accompanied by performance standards that define multiple degrees of mastery of those content standards (e.g., basic, proficient, advanced). These performance standards will help students, parents, and teachers understand how well students are doing in relation to the content standards and may motivate students to go beyond "passing" or "failing." They would also help schools and districts target resources to those students in most need of additional support.

CRITERION 2

Content standards should reflect the strengths of the relevant academic disciplines.

The academic disciplines represent the traditional way of organizing the knowledge and forms of thinking central to education's mission. Content standards should clearly show how they build upon the strengths of the core academic disciplines, such as mathematics, English, and science.

The conceptual structure and coherence of the academic disciplines reinforce and extend their educational significance. For example, the student with a general familiarity with a particular discipline can move more readily from the material in one area of that discipline to another. In addition, each discipline offers methods, rules, and strategies for solving problems and finding answers that fortify the student's ability to pursue new knowledge in that discipline.

One aspect of the quality of content standards depends on their being grounded in the core academic disciplines. The most straightforward way to ensure this is simply to develop and organize content standards by subject area. This approach enables standards to directly reflect the positive characteristics of discipline-based knowledge and promises to preserve the integrity of the established ways of acquiring and using knowledge represented by the disciplines.

At the same time, the academic disciplines are not wholly separate from one another, and content standards can be designed to reflect or encourage the integration of the ideas and activities of related disciplines. For example, mathematical ideas and procedures play a large role in scientific thinking.

Schooling can arrange to group academic disciplines together based on their affinities to each other. Primary schools often do not differentiate staff by discipline, and teachers are responsible for teaching several or all of the core disciplines. Even when schooling is devoted to learning in a specific subject, teachers can draw the relevance of knowledge and ideas from other disciplines into the lesson at hand.

It is essential that content standards integrate the academic disciplines to serve a meaningful and intellectually sound educational objective. Standards may be written across the various disciplines in order to emphasize their commonalities or to highlight opportunities to combine the curricula of several disciplines into single learning experiences and to apply them to "real world" problems. The intellectual and practical advantages of such innovative approaches make them worth trying, so long as those using them keep in mind the importance of preserving intellectual coherence, rigor, and academic excellence.

CRITERION 3

Content standards should be specific enough to convey clearly the important academic knowledge and skills that all students should learn, but broad enough to allow for multiple approaches to curriculum, instruction, course design, and assessment.

Not only should content standards define both the knowledge and skills that are essential to understanding each academic subject, but also they should form the foundation for the other components of the education system. They should guide the development of curricula — including course syllabi, instructional materials, teaching activities — that facilitate students' progress toward meeting the content standards. In addition, content standards should form the basis for the development of performance standards and assessments so that students' progress toward achieving the learning goals embodied in the content standards can be measured.

Therefore, content standards should be specific enough to ensure that both curricula and assessments will focus on a common destination. At the same time, however, standards should not be so prescriptive that they inhibit the creativity and professional judgment of individual teachers or deny them the flexibility that they need to tailor instructional strategies to the individual needs of their students. Nor should content standards preclude individual districts and schools from supplementing the curriculum content to reflect their community's special traditions or interests. The balance between sufficient specificity and adequate flexibility is a delicate one, and that balance is likely to vary from jurisdiction to jurisdiction.

Differences in the specificity of content standards originate in the variety of the directives given to standards developers in states and districts. States that have a history of strong local control are likely to have less specific standards than states that provide strong direction to districts. District standards are likely to be more specific than state standards. The following considerations are important aspects of evaluating whether standards are written with appropriate specificity.

- ☑ *The impact of content standards on the curriculum.* One of the major benefits of establishing content standards is that they can lead to a certain uniformity in the knowledge and skills learned by students across the state or the district. Content standards can contribute to equity in a school system by helping to ensure that all students, regardless of their geographical location or the wealth of their community, receive a rich and useful body of knowledge and skills. Content standards can also help ensure a more seamless instruction

as students move through various levels or transfer from school to school. Content standards can bring more focus and integration to textbook design and professional development also.

Content standards should not impose demands on classroom teachers that are unrealistic or do not match the needs of their students, however. Content standards should not dictate particular instructional strategies or specific lesson plans. They should describe a core of important knowledge and skills in each subject area and leave the details to the professional judgment of educators. While standards can support specific curriculum designs, their use should not preclude innovative curriculum that addresses the same content in more effective ways.

In addition, the effort to achieve standards should not consume the entire school day or the entire school year. Instead, local communities and schools should be able to devote a significant portion of the curriculum to content that matches their individual interests and needs. Finally, standards should reflect a careful consideration of developmental appropriateness and cultural diversity.

✔ *The impact of content standards on assessment.* Another important benefit associated with content standards is that they can form the basis for effective assessments of what students are being taught in the everyday classroom. These assessments can inform educators and policy makers whether students are meeting the content goals set out for them, where instructional practices need to be improved or when additional resources need to be allocated. Parents can use the results to determine whether their children need extra help. Also, when tied to important events such as graduation, assessments can motivate students. Assessments based on content standards eliminate many of the inappropriate practices created by overreliance on commercial standardized tests, which are often divorced from the everyday curriculum and test only a narrow range of skills.

To make sure that assessments are linked to the content standards and the classroom curriculum, the content standards must be detailed enough to guide the test developers about the knowledge and skills that should be assessed. Also, it is important that the standards clearly articulate what students will be expected to know and be able to do so that parents, teachers, and students understand expectations.

However, content standards that are too detailed can lead to inappropriate assessments. For example, some content standards are so detailed that they are simply a list of facts and skills. Such standards

tend to lead almost inevitably to multiple choice tests of limited portions of content areas, instead of more complex assessments that require students to think and analyze, show their work, and write in a way that clearly demonstrates breadth and depth of knowledge and skills. It is essential for standards to establish an intellectual and conceptual framework for each subject area, but there is a limit to the number of individual facts and skills that standards should enumerate.

✔ *Usability at the intended levels and in the intended contexts* There is a long tradition of local control in America's education that remains strong today. The extent to which a state or district should dictate what students in all the schools within its boundaries should learn is a question that will have to be resolved in each jurisdiction. In some areas, the public and the education community may be comfortable with decisions made at the state level, while other areas may wish to leave more decisions at the district or school level. However, there are positive and negative aspects of either approach.

Content standards do not stand alone. Many teachers and parents will want more information about expectations of students than content standards provide. Curricula based on the content standards are important means to fill in the detail that standards lack. Examples of curriculum "building blocks" can be incorporated into the content standards presentation to help readers visualize ways that standards may be achieved. Some states develop curriculum frameworks and instructional resources in addition to standards in order to offer more specific information about what to teach and how to teach it.

CRITERION 4

In addition to content standards themselves, standards setters should establish a plan to implement the content standards, through development of performance standards, assessments, and accountability measures. In the interim, content standards should be illustrated by examples of student work that meet the content standards.

As discussed earlier, content standards are the foundation, but are only one part, of a systemic reform package. States and districts should also plan to develop assessments, which measure student learning against the standards, and performance standards, which articulate in detail the levels of performance that students are expected to demonstrate on the assessments.

While content standards indicate at a general level the knowledge and skills that students will be taught, performance standards provide concrete information about what exactly students will be expected to learn. A content standard, for example, may state that students are expected to learn about the causes of the Civil War. A performance standard would describe a satisfactory examination answer on that subject.

Developing performance standards is a long and highly iterative process. Performance standards are closely connected with assessments; ultimately, they will be articulated in terms of student responses to tasks on assessments. For example, Maryland and Kentucky have established performance levels on their statewide assessments. In order to generate this type of performance standard, performances on specific assessment items must to be examined and related to the content standards. Consequently, meaningful performance standards can be developed only as related assessments are developed; they cannot be developed contemporaneously with content standards.

Nevertheless, high quality content standards should contain some illustrations of performance that will give readers a better sense of what is expected of students while the assessments and performance standards are being developed. For example, a content standard in English language arts pertaining to essay writing may include a sample student essay. The exemplar should also be accompanied by an explanation of how it meets the standard.

Standards should be dynamic; states and districts will have to undertake almost continuous review and evaluation of their standards as they learn more by implementing those standards at the school and classroom level and as content in each of the subject areas grows and evolves. Certainly, the development of assessments and performance standards, as well as curriculum, is likely to lead to a reexamination of content standards with subsequent clarification or revision.

CRITERION 5

Content standards should be world class standards.

In order for the United States to remain internationally competitive and for American families to maintain and improve their standard of living, businesses must be able to hire youngsters with the knowledge and skills necessary to compete in the global economy. If students in other countries are outperforming our own, perhaps those nations are doing a bet-

ter job of developing the potential of their young people. By looking overseas, the United States may aim higher when judging the potential of its young people.

But how does one determine whether content standards are "world class"? The issues that must be addressed in answering that question include the following:

- ✔ *With what precisely does one compare the standards?* A usable definition of the term *world class* makes it necessary to specify what exactly will be compared. One approach is simply to compare the content standards with curriculum documents from other countries. Another is to compare not only the content standards, but also the entire school system, including overall structure, assessment methods, student transition points, performance levels, and the portion of the cohort meeting those performance levels. Perhaps the most valid method (though clearly a difficult and expensive one) would be to administer assessments based on the content standards both to U.S. students and to students in other countries and then compare the performance of the groups.

- ✔ *What is an adequate comparison?* How many comparisons are needed. Is a comparison with one country enough? Could states simply compare themselves to another state that has already benchmarked itself at the world class level?

While a single comparison would be the simplest and most cost effective choice, it is unlikely that a single foreign system is the best in the world at everything. Some countries have strong mathematics programs; others may excel in elementary education. This reality is likely to necessitate multiple comparisons.

Currently, the models and resources for international comparisons are limited because work in this area is just getting under way. The International Education Study (IEA) and the Third International Mathematics and Science Study (TIMSS) provide some comparative data on student achievement, but do not directly address the comparability of standards and assessments across nations.

CRITERION 6

Content standards must be understandable and convincing to the lay public.

Standards need to be developed through a broad-based, consensus building process that involves the education of the community at large,

parents, the business community, and the broader public. Such a process will help ensure clear agreement on what all students need to know and be able to do. Everyone will be more likely to enthusiastically support the implementation of such standards.

In addition, teachers, parents, and communities must understand that content standards are a starting point for a concerted effort to improve students' achievement. Examples of student work, curriculum, instruction, and assessment are all effective means of illustrating the role of content standards in education improvement. All stakeholders must share the wider vision of systemic improvement from the early stages so that they can understand both the substance and implications of content standards. This understanding is essential in convincing the public of the need to support and implement the standards.

Questions to Guide Analysis

CRITERION 1
Standards should expect and support all students achieving to high levels.

- ✔ Are all students given the opportunity to reach high standards as evidenced by availability of
  - ✔ Curricular and instructional resources related to the standards?
  - ✔ Substantive, ongoing professional development related to the standards?
  - ✔ Rigorous certification of teachers and administrators related to the standards?
  - ✔ Guidelines for reporting adequate yearly progress for Title 1?
- ✔ Do the standards reflect high levels of achievement/ knowledge in the disciplines?
  - ✔ Have they been compared to national standards and those of other states?
  - ✔ Do they promote a curriculum and assessments that go beyond the current status quo?
- ✔ Can students demonstrate achievement in a variety of ways as evidenced by
  - ✔ Clear, measurable performance standards?

- ☑ Transitional assessments that are related to the standards, allow for diversity in students, and provide for comprehensible reporting?
- ☑ Plans for newer assessments that are aligned to standards, support high achievement, and reflect sound instruction?

CRITERION 2

Content standards should reflect the strengths of the relevant academic disciplines.

- ☑ Is each discipline covered adequately/ completely/ appropriately?
- ☑ Are the standards linked across the disciplines (common language, common framework, cross references)?
- ☑ Do the standards reflect knowledge of teaching and learning in the discipline(s)?

CRITERION 3

Content standards should be specific enough to clearly convey the important academic knowledge and skills that all students should learn, but broad enough to allow for multiple approaches to curriculum, instruction, course design, and assessment.

- ☑ Can the standards be used to guide reporting, assessment, and accountability?
- ☑ Do the standards reflect a balance between knowledge and skills?
- ☑ Is there evidence that curriculum guides can be developed and instructional decisions made based on the standards?
- ☑ Is there evidence that the standards allow for the degree of local flexibility that is characteristic of the state while providing clear, specific guidance?

CRITERION 4

In addition to content standards themselves, there should be a plan to implement the content standards, which would involve the development of performance standards, assessments, and accountability measures. In the interim, content standards should be illustrated by examples of student work that meet the content standards.

- ☑ Does the state have a plan for involving districts?
- ☑ Does the state have a plan for moving from standards to their implementation in instructional practice?

BEST COPY AVAILABLE

- ☑ Does the state have interim performance indicators and assessments?
- ☑ Does the state have a process for ensuring alignment?

CRITERION 5

Content standards should be world class standards.

- ☑ Have the standards been compared to standards in other countries, other states, other districts, and to national standards?
- ☑ Do the standards promote the level of achievement reflected in assessments and performances of students from across the nation (NAEP), other states, and other countries?

CRITERION 6

Content standards must be understandable and convincing to the lay public.

- ☑ Are the standards clear, concise, and jargon-free?
- ☑ How has the public been involved in the process of developing and approving the standards? Have there been responses to various contingents of the public, or have a vocal few held sway?
- ☑ What is being done to continue to engage the public as standards are implemented?
- ☑ How have teachers been involved in planning the implementation of standards?
- ☑ What evidence is there that teachers and the public are being convinced that the standards are to improve teaching and learning for all children?

OVERALL CONTEXT

- ☑ What was the overall context in which standards were developed?
- ☑ What is the stated purpose of the standards?
- ☑ Who sanctions them?

BEST COPY AVAILABLE

The following are standards criteria as set forth by existing documents, education organizations, or initiatives.

American Federation of Teachers

- ☑ Focus on academics
- ☑ Grounded in core disciplines
- ☑ Specific enough to assure the development of common core curriculum levels
- ☑ Manageable given the constraints of time
- ☑ Rigorous and world class
- ☑ Include performance standards
- ☑ Include multiple performance levels
- ☑ Combine knowledge and skills equally
- ☑ Not dictate how the material should be taught
- ☑ Be written clearly enough for all stakeholders to understand

National Alliance of Business

- ☑ Framework of common principles
- ☑ All students have opportunity to master academic subject matter held to world standards
- ☑ One set of standards for all students
- ☑ Have a common core of skills
- ☑ Reflect "real world" requirements
- ☑ Voluntary
- ☑ Dynamic
- ☑ Include criteria against which to measure performance
- ☑ Business must be involved
- ☑ General public must be able to understand
- ☑ Make workplace performance requirements part of subject matter standards
- ☑ Use skills and abilities from SCANS
- ☑ Accountability

BEST COPY AVAILABLE

Council of Chief State School Officers

- ☑ Challenging for all students
- ☑ Open, public process
- ☑ Indications of achievement at the elementary, middle, and secondary level
- ☑ Include content and performance standards together
- ☑ Assessable
- ☑ Achievable
- ☑ Part of reform strategy
- ☑ Internationally competitive
- ☑ Research based
- ☑ Important and focused
- ☑ Useful to learner
- ☑ Inspirational to students and teachers
- ☑ Usable, adaptable, and cumulatively feasible over time
- ☑ Balanced

Higher Education Advisory Group on K-12 Standards

- ☑ Curriculum standards encourage a set of core intellectual skills
- ☑ Ongoing mechanism to agree upon, specify, and explain nature of core intellectual skills
- ☑ Higher education should help develop nature and content of standards and explain their importance to public
- ☑ Performance standards have multiple levels
- ☑ Content standards allow variation among curricula course work
- ☑ Content and performance standards provide evidence of core intellectual skills
- ☑ States incorporate standards into teacher training

BEST COPY AVAILABLE

National Education Goals Panel Report

- ☑ World class
- ☑ Important and focused
- ☑ Useful in developing citizenship, employment, and lifelong learning
- ☑ Reflective of broad consensus building
- ☑ Balanced
- ☑ Accurate and sound
- ☑ Clear and usable by parents, teachers, and students
- ☑ Assessable
- ☑ Adaptable
- ☑ Developmentally appropriate
- ☑ Feasible
- ☑ Encouraging ability to integrate and apply knowledge and skills

BEST COPY AVAILABLE

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE
COUNCIL OF CHIEF STATE
SCHOOL OFFICERS

INTRODUCTION

5

IMPLEMENTATION MODELS AND PROTOTYPES

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

POLICY

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

BEST COPY AVAILABLE

POLICY

Education and policy leaders face many challenges while moving from standards development to standards implementation. Previous sections of this Tool Kit address challenges such as what might be included in content standards, instructional and assessment products and activities that are integral to standards implementation, and criteria for evaluating standards. This section addresses two key challenges on the policy side of standards development and implementation. The first challenge is to establish the systemic programs that are essential to standards-based reform. The second challenge is to devise a strategy to ensure a smooth roll-out of standards.

There are several ways this section can assist state or district efforts:

- ✔ Legislators and legislative staff can use this section to determine which policies support standards-based school reform, as well as which policies impede the progress of education reform.
- ✔ Policy makers can use this section to determine which state programs must accompany content standards.
- ✔ State or district leaders can use this section to strategize for the process of rolling out standards.

This section is focused on emphasizing that standards development is only a part of a larger reform agenda. The term "standards-based reform" denotes the systemic quality of a reform program that is aimed at the development of an educational experience that will enable individuals to seek and manage necessary information, think analytically and creatively, communicate effectively with others, and solve novel and difficult problems—the capabilities needed for success in our modern society and economy. The particular policies a state must develop to promote this vision of learning will vary from one context to the next. Therefore, this section does not attempt to describe specific policies; rather, this section outlines the important programmatic and strategic considerations around which policies must be developed.[1]

Challenge One: Systemic Programs

It is important to remember that in most education reform efforts, standards are a necessary piece of the reform agenda, but are only a piece.

[1]For more information, see the State Education Improvement Partnership's publications *From the Summit to the States: Linking Standards to Learning and Cross-Cutting Issues of Standards-Based Education Reform.*

While it is likely that the development and implementation of standards will have some educational benefit, it is rarely expected that standards alone will lead to widespread improvement in teaching and learning. Rather, standards are a foundation upon which other improvements are built. Rigorous standards provide a vision of deep understanding on the part of all students. For this vision to become reality, five key program areas are critical: the alignment of standards with curriculum and pedagogy; professional development; assessment and accountability systems; local capacity building; and collaboration.

ALIGNMENT OF STANDARDS WITH CURRICULUM AND PEDAGOGY

The vision of standards-based reform is an entire system geared toward high performance from all students. Standards call for a deep understanding of complex curriculum, and will be useful only to the extent they are reflected in classroom practice. It is not reasonable to expect deep understanding of content from students without offering them opportunities to deal with complicated issues over time. Unfortunately, many teachers do not have access to state or national standards, let alone have standards connected in concrete ways to their instruction. Further, most new standards are built on an assumption of a dramatic change in pedagogy. Teaching students to use facts to solve a problem, for example, requires different teaching skills than does teaching students to remember discrete facts. Therefore, it is not reasonable to expect teachers to be able to teach to the standards without offering them new curriculum and pedagogy.

For these reasons, curriculum experts are needed to support standards implementation by developing curriculum tied directly to the standards. However, this need comes at a time when state agencies face a diminished capacity for precisely this type of work. Due largely to budget cuts, state curriculum specialists have been forced to become generalists or have been eliminated altogether. Expecting deep understanding from students is not reasonable if curriculum developers do not have that deep understanding, or are not able to construct curricula for all subject areas and all grades due to limited staffing funds.

In an effort to deal with this need in the face of limited state agency staffing, some states are beginning to enlist professional curriculum organizations. State policy should support the working relationships between standards developers, professional curriculum organizations, and state agency staff, and encourage further research on teaching for understanding and student performance.

PROFESSIONAL DEVELOPMENT

Resources, ideas, and commitment must come from all levels to support the prodigious changes in teaching and learning that are expected by standards. As noted above, teachers are being told not to continue using current methods, but to implement improved curriculum and pedagogy. As would be expected with the implementation of any program, current teachers will need time to learn, review, and further develop the curriculum based on new standards. They need unprecedented support to develop deep competencies in their disciplines and to incorporate new teaching styles into their classroom strategies. Teacher education programs must develop new approaches to pre-service education. These changes require more resources and new approaches than ever marshaled before.

Many means can be used to promote the involvement of state and local education agencies, institutions of higher education, local teacher organizations, administrators, and professional development providers in teacher change. The most obvious is state funding for teacher professional development, which must be increased and directed toward the standards. Continuing development activities must be dynamic, as the standards will not be static. A second includes licensure and certification requirements, which influence both pre-service teacher education and in-service professional development. Certification requirements seem to be the best opportunity that state policy makers have to influence institutions of higher education to change their programs to come into alignment with the standards. In addition, recertification requirements and teacher competency tests guarantee that current teachers are keeping up with developments in the field. A third is the infusion of technologies into a school not only to support instruction, but also to bring together teaching professionals and to create long term, subject-based support networks. A fourth is the institution of site-based management and other programs that support the reorganization of the classroom and the school. A fifth uses findings from pilot projects to raise expectations for each school and to show what can be done and how that can be accomplished.

Teacher change has become one of the most critical pieces of education reform in most states. Students cannot be expected to achieve at high levels if teachers are not prepared to help them. Stringent quality requirements and consequences for teachers must be in place if those quality measures are to be achieved. When teachers are given the necessary resources, high stakes requirements can be used to continue to advance the teaching profession.

ASSESSMENT AND ACCOUNTABILITY SYSTEMS

Once common visions, consistent guidelines, and essential services and support are in place, each part of the system can legitimately be held accountable for fulfilling its role in providing effective education to all students. All stakeholders, including teachers, parents, community members, and elected officials will want the system to serve students in the most efficient and effective way possible.

An accountability system must be carefully designed and implemented. States face the challenge of determining what counts as an indicator of education reform and deciding how to handle that information once it is measured. It is rarely evident what information is needed in order to hold the education system accountable. Even more problematic, it is difficult to determine exactly for what the various parts of the education system can be held accountable. Reliable indicators of educational opportunity are difficult to develop, measure, and attribute to any one part of the system. For example, if a group of students does very well on a test, can this be attributed to the teacher, the principal, or the community's economic context? If a high percentage of students graduate from high school, can this be attributed to the superintendent, the school board, or the fact that a high percentage of students' parents graduated? In this age of accountability for results, nearly every state compiles and distributes report cards for the districts and schools. But just how much can a report card tell a stakeholder about a reform program?

Almost every standards-based indicator system will have as its cornerstone a student testing system. After all, student achievement is the ultimate goal of standards-based reform. The area of student assessments is one of the richest and most enigmatic areas of standards-based reform. Developing a valid, reliable, easily interpreted test is difficult. First, new types of learning are expected from many reform efforts, leading to a need for new tests of learning. Such tests are still being developed and tested. Second, education reforms take time to implement and even more time to lead to changes in student performance, yet there is often an expectation that student assessments will tell a story about a reform's success immediately. It is important to determine the appropriate timing for student testing. Third, expected outcomes of student assessments vary. The way tests are developed and results are reported differs depending upon whether the purpose of the assessment is to serve as a diagnostic tool, a tool for rewards and sanctions, or a high stakes assessment for individual students. Finally, as stated above, it is not clear to whom one can attribute good or bad test scores. Because student assessments are the cornerstone for an accountability system, it

is important that assessments be considered at the same time a state is developing and implementing a standards-based reform program. For example, in order to address the relationship between teaching and testing, the state of Maryland involved both curriculum and assessment professionals in the development process. Through collaboration, both test and curriculum developers gained a clear understanding of the reform program's learning goals. A new curriculum was designed to foster these goals and a new assessment was created to measure them.

It is also important to remember that student assessments are not the sole form of accountability data. No one assessment tool can provide all of the information needed by various stakeholders in the system. States and localities need an indicator system that will enable decision makers to measure and report progress toward achieving goals; understand the characteristics of the student population; guide improvement, reward success, and take action when success is lacking; and determine the best use of limited resources to improve student learning.[2] It is a challenge to compile and report performance indicator data so that all stakeholders can read, understand, and use the information. A "good" set of performance indicators should be built upon the state's goals, standards, and objectives.

In addition, the purpose of an accountability mechanism must be clear in order to appropriately use the information produced. Some accountability systems have been focused on rewards and sanctions, while others have been focused on diagnosis and correction. Tests can be used to determine whether teachers should be certified or tenured, whether a particular program is working, to diagnose problems a school faces, and to determine whether or not an individual student will graduate. It is important to consider and review the stakes involved in an accountability and assessment program with attention to its uses and results and to be prepared to make modifications if the intended purposes are not being met or are counter to the ultimate reform goals.

Finally, whether an accountability system focuses on sanctions or diagnosis, it should include mechanisms for support such as service, advice, and resources to school districts and schools. With such extra support, a school that does not meet the state's ambitious requirements can grow to meet them, which is the ultimate goal of accountability.

[2]For more information see the State Education Improvement Program's publication *Measuring Results: Overview of Performance Indicators*.

BEST COPY AVAILABLE

LOCAL CAPACITY BUILDING

Teachers are not the only educators being required to change. Superintendents, principals, state and local education agency personnel, and other child-serving agency staff are also integral members of the reform effort. Schools are beginning to serve as a locus for responsibility, which implies profound changes in school management and accountability. Such changes will require both fiscal and organizational support from the state level, as school leaders are given the training and professional autonomy needed to lead local reform implementation.

For example, local administrators must ensure that school policies give teachers the necessary professional autonomy and support, while at the same time involve parents and community members in decision making. Previous bureaucratic approaches to administration are being replaced by quality management approaches that build on a common vision and the involvement of every member of an organization in improving results. Lessons from the business world indicate that this approach to management is successful, but requires training.

While local management approaches must change, so must the regulatory role of state and local education agencies. Regulation alone rarely leads to behavioral changes. Successful standards-based reform requires state and local education agency staff to be co-partners with each other and with schools, classrooms, teachers, and students. To serve as "good co-partners," the agencies must learn with the schools what their needs are, what reforms are planned for implementation, and how to provide services in conjunction with the school's and/or district's needs such as research and sustained technical assistance.

All of these changes in roles and responsibilities will require an investment in more than just teacher professional development. Standards based reform requires that the capacity of all local participants be considered and that support be given to increasing that capacity.

COLLABORATION

The Council of Chief State School Officers' policy statement on systemic reform discusses the need for collaboration among all child-serving agencies.

> The central purpose of schools is teaching and learning, but we know that the conditions in which children live can affect their readiness and ability to learn. Although schools alone cannot meet the myriad developmental needs of young peo-

> ple or the human service needs of their families, they have an interest in seeing that these needs are met in order to achieve the primary goal of schools. Students who need additional support services often receive assistance through several private and public agencies or organizations. In many cases the work of these agencies is not coordinated, and fails to serve the "whole child." We need to establish clear, strong connections between these social services and the schools. Such connections can range from frequent adult contact and guidance within schools, to encouragement of more stimulating nonschool opportunities for students, to fostering connections between vulnerable children and their families and health and social supports. This is not to say that schools should assume these extracurricular functions. Rather, the education system must help to create more supportive environments through collaborative programming between schools and other child-serving agencies.

Education leaders have little control over the goals, functions, and approaches of non-education agencies and personnel. However, state policy makers can confirm that similar reforms are taking place in other program areas in addition to education. Such reforms will have an additive effect on student performance, as the whole child is served in a coherent manner.

Challenge Two: Strategies

As states begin the process of implementing standards at the classroom level, they must have a strategy for maintaining continuity in the reform effort and providing incentives for improvement and consequences for lack of improvement. Strategies will vary depending upon the historical, political, social, and economic context in a state in which education reform is attempted. However, there are key issues that seem to cut across the states, including the need to involve local stakeholders, to revisit the standards debate at the local level, to create partnerships among leaders, and to embody components of standards-based reform in legislation.

INVOLVEMENT OF LOCAL STAKEHOLDERS

Many states have included a wide range of state- and national-level curriculum experts, education agency personnel, and elected officials in the standards implementation process. It is essential that local stakeholders

such as teachers, principals, mayors, and parents be involved early in the process. Not only does standards-based reform rely on the acceptance by local stakeholders, but those local stakeholders will also bring valuable knowledge to the development and implementation process. Few people know better what it will mean in the classroom and the community to push for high standards and to implement particular reforms than those currently immersed in the local setting. They will be the ones most directly affected by changes related to standards and will have to interact with the students and families who are also affected by the standards-based reforms. Local stakeholders will be held accountable for results after implementation of the standards and will, therefore, be concerned about the nature, content, and purpose of the standards. State leaders must be attentive to local concerns by involving local representatives in early development and implementation activities. Amending reform programs as the need arises ensures that local needs and contexts are considered.

State leaders must be aware that if local stakeholders do not understand, accept, or care about the standards, the standards and any related improvements may be implemented differently than intended by the standards development teams; they may founder because of half-hearted attempts at implementation; or they may be derailed because of political resistance. State leaders must do what they can to provide resources, staff, and general support to get these important efforts off the ground. Resources such as local media, electronic platforms such as e-mail and the Internet, and professional and special interest group newsletters and events can be used to disseminate information and questions to broad audiences.

These efforts will allow problems with reforms to be aired and dealt with early in the process, preventing these problems from derailing efforts after a great deal of time and energy has been expended. A concerted effort to engage local stakeholders will ensure that reforms take into account local considerations and that stakeholders possess an understanding of the underlying goals of reforms, thereby further confirming that implementation efforts will be congruent with reform goals.

REVISITING THE STANDARDS DEBATE

Regardless of the extent to which local constituencies have been involved in the state standards development efforts, discussions about the use of standards must take place again on the local level. Each phase of the standards roll-out will require that the same policy issues be reviewed by a new and expanding base of stakeholders. Difficult questions that have

been addressed in multiple meetings will be raised again and again. Ideas taken for granted by those involved in standards development will be new to local audiences. Leaders must make the case for standards over and over as each new teacher, school, or community feels the effects of standards-based reform. Local stakeholders will want to know how standards will affect their children or children in their classrooms. Academic jargon or erudite testimonials about standards are not needed. Rather, a straightforward, honest sense of how standards will support education improvements in their schools is needed.

The questioning of the standards by local communities may be viewed as a challenge of the work of those who have invested a great deal of time, effort, and care in developing the standards. Preparing the individuals who have been active in the standards development process for the next wave of debate and critique will provide timely resolutions. This process can be viewed as a type of formative evaluation. The roll-out of standards will offer a fresh perspective to people who have been close to the standards over a long period of time. Making the case in new venues and for new audiences can actually support the process of diagnosing and fine tuning standards.

Each state will likely have different issues arise related to making the case for standards, based on unique political cultures and histories. However, there are some key issues that seem to resonate in discussions about standards across the states. When making the case for standards, stress their role in creating a system-wide vision, articulating what students should know and be able to do, supporting accountability efforts, and fueling the quest for excellence and equity. Following are some positions that can be used in local implementation efforts.

- ✔ *Standards provide a system-wide vision for higher achievement and deep understanding.* Standards provide an explicit vision of the substantive purposes and goals of the system. The vision of the standards movement is one of deep understanding. Education is often held out as a way to prepare for changing economic and social conditions. To thrive in the expanding service sector and the "information age," people must possess skills and abilities. It is the hope of the standards movement that formal education will help develop students who can seek and manage necessary information, think critically and creatively, communicate effectively with others, and solve novel and difficult problems. By stating these goals explicitly, standards provide coherence to schooling. Standards set the target; all education resources, policies, and practices can be aimed at supporting student achievement to reach that target.

- ✔ *Standards tell what students should know and be able to do:* Standards provide substantive criteria for accomplishment. Traditionally, expectations of students have not been clear. Achievement of individual students has been measured with respect to achievement of their peers; what children know and can do with respect to specific, substantive learning goals has not been measured. Standards provide a measuring stick against which the success of each student in reaching common, high expectations in various content areas can be determined. Often parents have not been informed about how to help their children attain the skills, knowledge, and abilities they need to succeed in school. Standards provide parents a rather detailed map of the schooling process and help them to support their children's achievement.

- ✔ *Standards help stakeholders hold the system accountable:* The measuring stick that standards provides will help parents, teachers, and community members evaluate the effectiveness of their schools. It has been traditionally very difficult for parents to know whether or not the schools were providing their children with high quality educational opportunities. State- or district-wide standards tell parents to what exactly they can expect their children will be exposed in schools. Knowing this, parents can determine whether or not their children are being well served by the education system.

- ✔ *Standards promote both equity and excellence:* A major development in educational theory and policy has been the realization and articulation that all students can achieve to high standards, given the opportunity and expectation to do so. Standards provide the expectation. Using the goals that are explicitly spelled out in the standards, students, families, and communities are able to demand that schools provide the opportunity. Holding all students to the same high standards will guarantee that the system is focused on excellence from and for all students.

CREATING PARTNERSHIPS AMONG LEADERS

A standards-based reform effort can be insufficiently narrow or ineffectively wide. It is important that key leaders agree on the scope of a reform plan and work together toward the goals of the reforms. There are numerous activities that the various stakeholders can prioritize, and these will vary from state to state. Following are a few examples:

- ✔ Governors have unequaled access to agenda-setting opportunities and can use those opportunities to push for education. They can provide leadership and vision for education improvements; give credibility to standards-based reforms and keep the public's attention focused on them; and build broad coalitions in support of education reform among business, parents, and other community members.

- ✔ State legislators can discuss and enact reform legislation; require education agencies to demonstrate their reform plans and achievements; hold governors accountable for commitments; and see that actions transcend political rhetoric.

- ✔ State boards of education can lead the nitty-gritty work of standards development and implementation and make sure that local boards of education have the needed information and supports to carry out reform efforts.

- ✔ State education agencies can offer expertise on education issues; grant local districts and schools flexibility in the adoption and implementation of standards-based reforms; build the capacity of local education personnel; and conduct or compile research on key education issues.

- ✔ Mayors, local school board members, and other leaders can support the standards development process by bringing the local perspective to the process; providing state-wide efforts with the needed flexibility to conform to local contexts; and engaging parents and other community members in local reform planning.

- ✔ Superintendents, principals, teachers, and other local educators can develop plans to carry out reforms that are compatible with the overall systemic goals; become involved themselves in the standards-writing process; offer high quality professional development opportunities in their schools; and bring the public into the schools to encourage their feedback and support.

- ✔ Parents and community members can hold schools accountable for results and state leaders accountable for local needs; serve on site-based management teams; and send healthy, happy children to school.

- ✔ Business leaders can educate themselves about the type of workers they will need in the coming decades; keep their sights trained at high performance workplaces; and promote state and local reform efforts that build toward the 21st century.

Whatever the activity, it is clear that these efforts will work best when they are coordinated. Such coordination does not require a large expenditure of time and resources. If communication lines are created and maintained, most activities can enjoy the input of each and every leader.

EMBODYING REFORMS IN LEGISLATION

Education reform is a complex, continuous process. Changing leadership is inevitable in such a long term effort. Many states have embodied standards-related policies and programs in legislation in an effort to carry out reforms in a coherent manner. In 1991, the Oregon legislative assembly enacted a comprehensive piece of legislation entitled the *Oregon Educational Act for the 21st Century*. This Act outlines the characteristics expected of the Oregon system of public elementary and secondary schools, including such things as the assumption that all students can learn. Given this assumption, the Act requires "high, specific skill and knowledge expectations appropriate to the students' assessed learning rates at all instructional levels."[3] The Act established a process for developing educational standards; the Certificates of Initial Mastery and Advanced Mastery; alternative learning environments and services for those students having difficulty obtaining the Certificate of Initial Mastery; an early childhood and academic professional technical programs; and partnerships among business, labor, and the educational community. The Act further determined who would take responsibility for reviewing and revising the common curriculum goals, performing site visits of each public school district, establishing an information system, and collecting data and producing district and school profiles. Self-evaluations are required of each school district and school in the Act and a 21st Century Schools Advisory Committee was enacted.

In another example, Colorado amended its statutes in 1993 to add sections on "standards-based education in public schools" to its Education Reform section. The new statutes require that a "system of standards-based education will serve as an anchor for education reform."[4] The statute includes statements of goals, the creation and empowerment of a state standards and assessments development and implementation council, time lines for the adoption of state model content standards and assessments, the establishment of a resource bank on standards and instruction, requirements for the adoption of state standards by district,

[3]Section 3, Number 2, Oregon Educational Act for the 21st Century, House Bill 3563, 1991.

[4]Part 4 of article 53 of title 22, Colorado Revised Statutes, 1988 Repl. Vol., as amended. 22-53-401, Legislative Declaration.

temporary waivers of regulatory requirements, establishment of and time lines for the Colorado student assessment program, authorization for the state board of education to sanction districts that have not adopted content standards, and a plan for implementation.

It is fair to say that the education reform movements in both Oregon and Colorado have been aided by legislation. The process of drafting, debating, and enacting the legislation was important for putting education reform high on the state agenda. After passage, not only does legislation describe what ought to be done, but it also provides timetables for action and authorizes positions and agencies to ensure these timetables are met. Further, it keeps education reform on the state legislative agenda by requiring reports to the legislature and providing an opportunity for community and education leaders to keep education high on the state agenda.

Key points to consider are that education reform legislation be specific enough to clearly state the goals of standards-based reform, but not so specific as to unnecessarily limit the work of standards developers or eliminate creativity and initiative on the local levels. Provisions for waivers or flexibility within education regulations seem to have advanced reform in many cases. Legislation should also clearly state who has responsibility for what. To the extent that reasonable time lines can be spelled out in legislation, states can better hold schools accountable for results. Time lines must be reasonable as education reforms can be difficult and improvements in student achievement as a result of those reforms will not be apparent immediately. States and schools have found that unrealistic expectations for measuring results impede the reform process. Reform legislation should set forth authority and time lines for the development of indicators of results, so that indicators are developed in a thoughtful, intelligent manner.

Regardless of the specific details contained in legislation, it is key to put education high on a state's agenda, to develop reform goals that are articulated and consistent, and to provide leadership and direction to the many individuals involved. Legislation seems to be an appropriate and useful tool in the standards-based education reform movement.

Meeting the Challenges

To assist states in meeting the challenges associated with standards-based reform, the key components such as program areas, stakeholders, and activities that comprise a successful standards-based reform effort

have been identified. A successful standards-based reform effort must not only involve these key components, but also must ensure that standards, professional development and capacity building, and assessment and accountability systems are aligned both with the standards and with all other components of the system. In addition, the success of standards-based reform has been furthered by legislative support from the state.

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

INTRODUCTION

IMPLEMENTATION MODELS AND PROTOTYPES

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

6

POLICY

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

ASSESSMENT GUIDELINES AND RESOURCES

- ☑ CRITERIA FOR ALIGNMENT OF EXPECTATIONS AND ASSESSMENTS IN MATHEMATICS AND SCIENCE EDUCATION ..... PAGE 1
- ☑ IMPLEMENTING THE ADEQUATE YEARLY PROGRESS PROVISIONS OF TITLE I IN THE IMPROVING AMERICA'S SCHOOLS ACT OF 1994 ..... PAGE 47
- ☑ HOW SHOULD THE COMPREHENSIVE ASSESSMENT SYSTEM BE DESIGNED? ..... PAGE 64
- ☑ GUIDELINES FOR THE DEVELOPMENT AND ADMINISTRATION OF PERFORMANCE ASSESSMENTS IN LARGE-SCALE ASSESSMENT PROGRAMS ..... PAGE 78

CHIEF STATE SCHOOL OFFICERS COUNCIL

CRITERIA FOR ALIGNMENT OF EXPECTATIONS AND ASSESSMENTS IN MATHEMATICS AND SCIENCE EDUCATION

by Norman L. Webb

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
WASHINGTON, DC

The research reported in this paper was supported by a cooperative agreement between the National Science Foundation and the University of Wisconsin-Madison (Cooperative Agreement No. RED-9452971). At UW-Madison, the National Institute for Science Education is housed in the Wisconsin Center for Education Research and is a collaborative effort of the College of Agricultural and Life Sciences, the School of Education, the College of Engineering, and the College of Letters and Science. The collaborative effort is also joined by the National Center for Improving Science Education, Washington, DC.

The research reported in this paper was also supported by the Council of Chief State School Officers under a research and evaluation project on State Curriculum Frameworks and Standards in Mathematics and Science Education (NSF Grant No. REC-9554462) in Washington, DC. Any opinions, findings, or conclusions are those of the author and do not necessarily reflect the view of the supporting agencies.

Abstract . . . 3
Introduction . . . 4
Importance of Alignment . . . 5
Alignment in Principle . . . 6
Alignment of Expectations and Assessments . . . 8
Alignment in Practice . . . 9
Interpretation of Alignment by States . . . 10
Methods for Aligning Documents . . . 12
Sequential Development . . . 12
Expert Review . . . 13
Document Analyses . . . 15
Quality Control . . . 18
Specific Criteria . . . 18
Content Focus . . . 19
Categorical Concurrence . . . 19
Depth of Knowledge Consistency . . . 20
Range of Knowledge Correspondence . . . 23
Structure of Knowledge Comparability . . . 25
Balance of Representation . . . 27
Dispositional Consonance . . . 29
Articulation Across Grades and Ages . . . 30
Cognitive Soundness Determined by Best Research and Understanding . . . 30
Cumulative Growth in Content Knowledge During Students' Schooling . . . 32
Equity and Fairness . . . 33
Pedagogical Implications . . . 35
Engagement of Students and Effective Classroom Practices . . . 36
Use of Technology, Materials, and Tools . . . 37
System Applicability . . . 38
Conclusions . . . 39
References . . . 41
Glossary . . . 45
Appendix . . . 46

Abstract

Alignment is central to current efforts of systemic and standards-based education reforms in mathematics and science. More than four-fifths of the states have content frameworks in place in mathematics or science. A large number of these have some form of a statewide assessment to measure student attainment of expectations given in the frameworks. These reforms are based, in part, on the premise that student outcomes will be improved through creating coherent systems of expectations and assessments. Expectations are major elements of educational policy on what students should know about mathematics and science and what they should be able to do with that knowledge. Assessments are major elements of educational policy used to measure student achievement and are classroom tools used by teachers. All assessments used within a state or district constitute an assessment system. The purpose of this monograph is to define criteria for judging the alignment between expectations and assessments.

States and districts were found to use three general approaches for judging the alignment among expectations and assessments. A number of states and districts develop documents in sequence, such as first standards, then curriculum frameworks, and then assessments. These documents are aligned because they were developed to be, with the previously developed document forming the blueprint for the next document. Other systems hire experts to review expectations and assessments. A third approach, used by the Third International Mathematics and Science Study, is to systematically analyze both expectations and assessments using a common metric. Specific criteria for judging the alignment, however, were missing from all three approaches.

Twelve criteria for judging alignment grouped into five general categories are described along with examples and levels of agreement. The five general categories are content focus, articulation across grades and ages, equity and fairness, pedagogical implications, and system applicability. The criteria were created to provide guidance to state or district officials on the important aspects of expectations and assessments that need to be considered in some detail to have a coherent system where the power of these policy documents converges to better support students' learning of important mathematics and science.

The criteria were developed with the input of an expert panel formed as a cooperative effort between the Council of Chief State School Officers and the National Institute for Science Education. The monograph was supported by the Council under a research and evaluation grant from the National Science Foundation on State Curriculum Frameworks and Standards in Mathematics and Science Education. A shorter version of this paper was published as an NISE *Brief*, January 1997.

Introduction

Many states and school districts are making concerted efforts to boost student achievement in mathematics and science. These are not efforts aimed at simple face lifts, but attempts to develop deep, lasting changes in how students learn these critical subjects. This monograph is directed towards those in states and districts who are working to improve student learning through creating coherent systems of expectations and assessments. Other audiences are those who study reform, make decisions about reform, and are affected by reform. The intention of this monograph is to help people think more clearly about the concept of alignment and to examine what is required for important system elements of expectations and assessments to converge.

Educators, notably through efforts spearheaded by national professional associations, increasingly recognize the need for major reform in K-12 mathematics and science curricula and are embracing a vision of ambitious content for all students. Making this vision a reality means encouraging "a far deeper and dynamic level of instructional decision making" (Baker, Freeman, & Clayton, 1991). This is not something that can be done simply by mandating new accountability measures. At the heart of these efforts to make deep changes in instruction is the concept of "alignment."

The major elements of an education system must work together to guide the process of helping students achieve higher levels of mathematical and scientific understanding. Educators increasingly recognize that if policy elements are not aligned, the system will be fragmented, will send mixed messages, and will be less effective (CPRE, 1991; Newmann, 1993). For example, the Statewide Systemic Initiatives (SSI) program of the National Science Foundation is directed toward states, districts, and regions setting ambitious goals for student learning, through a coherent policy system established, in part, on assessments aligned with those goals. The Improving America's Schools Act explicated how assessments are to relate to standards: ". . . such assessments (high quality, yearly students assessments) shall . . . be aligned with the State's challenging content and student performance standards and provide coherent information about student attainment of such standards . . ." (U.S. Congress, 1994, p. 8). The U.S. Department of Education's explanation of the Goals 2000: Educate America Act and the Elementary and Secondary Education Act (which includes Title I) indicated alignment

of curriculum, instruction, professional development, and assessments as a key performance indicator for states, districts, and schools striving to meet challenging standards. As more and more weight is given to accountability throughout education systems, alignment between assessments and expectations becomes not only critical, but also essential.

Importance of Alignment

Assuring the alignment between expectations and assessments can strengthen an education system in important ways. Teachers give more credence to documents they understand are in agreement, are useful, and will serve to benefit their students. Teachers, already overloaded with responsibilities, are better able to attend to expectations and assessments if they provide a consistent message and have credibility. District curriculum guides, large-scale assessments, standards, state frameworks, and other policy documents overwhelm teachers. Often teachers' only strategy for coping with this barrage of directions is to ignore them (Cohen, 1990). Carefully aligned assessments and expectations, with input from teachers and others, add to the value teachers give to these documents and their willingness to make sense of these documents. Teachers are more apt to attend to student outcomes listed in expectations if they know these outcomes will be assessed by instruments that will give them feedback on how well their students did on these outcomes. Through understanding the link between expectations and assessment, teachers are more likely to find ways to translate what is being advanced by these documents into their daily work with students.

Aligning assessments with expectations can improve the efficiency and effectiveness of the education system. An aligned system can more effectively set priorities and allocate limited resources. Both expectations and assessments are important statements of what the system believes students should know and do. Better aligned goals and measures of attainment of these goals will increase the likelihood that multiple components of any district or state education system are working towards the same ends. Most students in their K-12 school career will have a number of science and mathematics teachers, maybe as many as 18 or more. Aligning goals carefully with the assessment system — all forms of gathering information about students' learning — is an important tool for mapping students' learning progress as they proceed through the system, responsibly allocating curriculum to teachers, and verifying when students' knowledge of important ideas is assessed. Not all learning can be assessed by large-scale assessments. Teachers are in a much better position to assess important learning

such as how well a student is able to perform a scientific inquiry or devise a mathematical proof. Aligning the assessment system with expectations serves as an inventory to help assure all outcomes are being assessed in some way. That is, the expectations are covered by the assessments. A careful analysis of alignment between expectations and assessments also will help reduce unnecessary repetition in the assessment system caused by over-assessing a few outcomes at the expense of ignoring others. Other important system functions, such as professional development, soliciting public support, and textbook selection, can be more effectively planned if the goals and the measurement of those goals are in agreement. A formal process for assuring agreement of policy elements and the verification of their agreement both are means for dealing with the complexity within the system and marshaling the support of others.

A formal process for assuring assessments and expectations are in alignment provides a response to those who challenge the system. Both state education departments and school districts face public scrutiny and review. In some instances, legal suits have been waged against education systems for not adequately educating students. A system will be in a better position to respond to any challenges if it has a well-defined and thoughtful process of assuring that it is accountable. Part of this assurance comes from confirming that assessments and expectations are aligned. That is, students are being assessed on what they are expected to know. Having a system where assessments are aligned with expectations, however, will not prevent challenges to the system, nor will it eliminate the need to validate the quality of the expectations. A formal alignment process employed by a district or state is one indication that these systems are assuming responsibility for assuring that students are learning what is expressed as important knowledge in standards, frameworks, or other statements of expectations.

Alignment in Principle

Two or more system components are aligned if they are in agreement or match each other. In the past, the most common educational use of the concept of alignment referred to the match between an assessment instrument (or instruments) and a curriculum. Here, alignment is analogous to instructional or curricular validity of a test (Harmon, 1991). A legal ruling in the 1981 Florida case *Debra P. vs. Turlington* emphasized the importance of assuring agreement between a curriculum and tests. According to this ruling, for a test to be fair, both curriculum and instruction must match the content coverage of the test (Madaus, 1983).

Legally, high stakes tests need to be fair by being "aligned" with curriculum and instruction.

Alignment has been used as a comparison between the psychometric qualities of an assessment and instructional uses of results. Baker, Freeman, and Clayton (1991) questioned the viability of using standardized norm-referenced tests to assess either the improvement of an individual's education or the impact of systemic education reform. Because traditional standardized test results provide only a ranking, not a level of performance, they obscure the meaning of test scores and are less aligned with instruction.

The form of an assessment can be as important as the content in judging alignment. "The content and form of an assessment task must be congruent with what is supposed to be measured" (NRC, 1996, p. 83). For example, an assessment using a short-answer format is not aligned with an intended purpose of measuring students' ability to frame questions for conducting scientific inquiry and to design an inquiry to address the questions.

Alignment does not only refer to a comparison between one assessment instrument with a curriculum, but extends to a set of assessment instruments or the assessment system. "The term *alignment* is often used to characterize the congruence that must exist between an assessment and the curriculum. Alignment should be looked at over time and across instruments" (MSEB, 1993, p. 123). A single assessment may not be well aligned with curriculum because it is too narrowly focused, but it may be part of a more comprehensive collection of assessments that is in full alignment with the curriculum.

With the advent of systemic reform and the increased prominence of standards, alignment is increasingly being used to characterize the agreement or match among a set of documents or multiple components of a state or district system. A model used for the evaluation of statewide systemic initiatives assumed that student attainment of high standards of learning requires significant improvements in classroom teaching, the use of more challenging curricula and materials, and the regular assessment of student learning integrated and aligned with instruction (Zucker, Shields, Adelman, & Powell, 1995, p. 1). Building on this more recent view of the match of important components within a system, alignment can be defined.

Alignment is the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do.

Alignment is intimately related to test "validity." However, important and useful distinctions can be drawn between the two concepts. Validity refers to the appropriateness of inferences made from information produced by an assessment (Cronbach, 1971). Alignment refers to how well all policy elements in a system work together to guide instruction and, ultimately, student learning. Of the many different types of validity (Messick, 1989, 1994; Moss, 1992), alignment corresponds most closely with content validity and consequential validity. For example, the degree to which a test is aligned with a curriculum framework may affect a test's validity for a single purpose, such as making decisions on the curriculum's effectiveness. But a test, or tests, and a curriculum framework that are in alignment will work together to communicate a common understanding of what students are to learn, to provide consistent implications for instruction, and to be fair for all students. In addition, tests aligned with a curriculum will be based on compatible and sound principles of cognitive development.

Alignment of Expectations and Assessments

This monograph is primarily confined to the discussion of alignment of expectations and assessments, two major elements of educational policy:

- ✔ *Expectations* of what students should know about mathematics and science and what they should be able to do with that knowledge. Expectations can be communicated in different ways. Educators can, for example, craft sets of standards or frameworks, ranging from broad vision statements to precise indications of expected performance and recommended instructional practices.

- ✔ *Assessments* used to gauge student achievement in science and mathematics and to indicate whether the expectations are being achieved. They can be used to formulate policy, monitor policy effects, enforce compliance with policies, demonstrate accountability, make comparisons, monitor progress toward goals, and/or make judgments about the effectiveness of particular programs. Assessments can refer to the collection or system of procedures used by teachers in classrooms and to district or state tests. Assessments include paper-and-pencil tests, but also encompass other forms of gathering information about students such as interviews and observations. The nature of assessments may be best represented by their development specifications along with instruments rather than by instruments alone.

Both expectations and assessments are now of great concern among educators and policy makers as key to standards-based education, systemic reform, and accountability (Chubin, in press; Ferrini-Mundy & Johnson, 1996; National Academy of Sciences, 1997). Because of the centrality of these two policy elements to current thinking on reform, this monograph is restricted to specifying ways of judging the agreement between only these two elements. There are, of course, many other important elements in any education system, such as professional development, instructional materials, college requirements, teacher certification, resource allocations, and state mandates. These elements cluster into four different strata: purpose, policy, programs, and practice (Bybee, 1995). How all of these serve to provide a coherent system is important.

The number of different components within a system suggest there are several ways to think about agreement among components. The type of alignment discussed in this monograph will be referred to as "horizontal alignment," meaning the degree to which standards, frameworks, and assessments work together within an education system and mainly at the policy level. This is different from "vertical alignment," which is the degree to which the elements among the strata in an education system (e.g., textbook content, classroom instruction, professional development, and student outcomes) are aligned with each other and with outside forces (e.g., national standards, public opinion, and work force needs). Horizontal alignment and vertical alignment will be used when necessary to distinguish between these two different directions of agreement.

Alignment in Practice

Judging alignment between expectations and assessments is difficult for several reasons. For one, both expectations and assessments frequently are expressed in multiple pieces or documents, making it difficult to assemble a complete picture. Also, it is difficult to establish a common language for describing different policy elements. The same term may have very different meanings when used to define a goal than when used to describe something that can be measured by assessment. For example, a viable goal in mathematics is for students to be able to use multiple strategies to solve problems. The intent of this goal is for students to have an assortment of strategies that they can draw upon to solve a problem. Sometimes this goal is measured by giving students a problem to solve and directing them to solve this problem in more than one way or explain their answer in more than one way. Solving a prob-

lem in more than one way does not fully assess whether students have a range of strategies for solving a variety of problems. Further, the policy environment in an education system can be constantly changing. New goals can be mandated, for example, while old forms of assessment are still in place. Ever-expanding content areas, advancing technology, and a growing body of research on learning also contribute to the complexity of identifying expectations and assessments.

Increasingly more states and districts are expressing expectations for student learning in frameworks or standards. At the end of 1994, 42 of the 50 states, the District of Columbia, and Puerto Rico had mathematics frameworks or curriculum-related documents (Blank & Pechman, 1995). About the same number had science frameworks. Ten of these states had combined mathematics-science frameworks. Nearly all of the other states were in the process of developing frameworks. Even though most of the frameworks were less than 10 years old, some states, such as California, have had frameworks for a number of years. The first version of California's *Mathematics Framework* was published in 1963 (Webb, 1993). Since then a revised framework had been produced by California about every 5-7 years as the first step in adopting each new round of K-8 textbooks. By spring 1996 40 states had content standards ready in mathematics and 38 had standards ready in science (CCSSO, 1996).

What constitutes a framework varies among the states. Some frameworks consist only of content standards while others go into more detail in specifying learning objectives and recommended instructional practices. Arkansas prepared a mathematics framework based on 13 content standards applicable to all grades; South Carolina defined mathematics content standards by grade ranges for each of 6 core strands; New Jersey specified 18 mathematics standards, including content, process, and learning environment standards for K-12. Science frameworks were as varied: Arizona defined 8 goals (K-12) that called for teaching higher order thinking skills and use of conceptual thinking; New Hampshire identified 6 curriculum strands, with curriculum standards for grades K-12 and proficiency standards for grades 6 and 10; Virginia outlined standards of learning objectives as well as descriptive statements that gave examples of the objectives; and Ohio provided local school districts with sample instructional performance objectives for four strands.

INTERPRETATION OF ALIGNMENT BY STATES

An analysis conducted in 1993-94 of the 25 states participating in the National Science Foundation's SSI program indicated that, even though many of the states had state assessments, a number of these states did

not have assessments aligned with the systemic reform goals for student learning (Laguarda, Breckenridge, Hightower, & Adelman, 1994). Many of these states were shifting their assessment policies. In science, the problem was not so much alignment as it was having any form of statewide assessment. Nearly all of the states had statewide assessments in mathematics, but fewer than two-thirds of the states had statewide assessments in science. SSI staff from nearly all states reported the availability of statewide assessment data judged to be appropriate for evaluating the SSI. However, when the SSI staff were asked to judge the alignment between the statewide assessments and the curriculum and instruction promoted by the SSI, fewer indicated there was an agreement. Only 10 of the 22 states with statewide assessments in mathematics and six of 16 states with statewide assessments in science reported content alignment based on the judgment of SSI staff.

In a fall 1995 annual survey conducted by the Council of Chief State School Officers, state assessment directors were asked, What does "alignment" mean in your state? In general, responses given to this question were very brief. The most common response indicated that assessment activities and content standards were aligned by design. For New Jersey, "Aligned means assessments will be based on the standards and indicators." The response from Michigan was that "the assessments are actually . . . designed to measure . . . outcomes and requirements stated in goals & objectives. Committees approve and reject items based upon their fit with goals and objectives." For North Carolina, alignment meant, "Curriculum frameworks provide the assessment framework for developing tests. All test questions, etc. are developed to meet the curriculum objectives." Some states selected existing tests to measure what was in their curriculum frameworks, but reported only a partial match. Although virtually all that was on the commercial tests used by North Dakota measured something in its frameworks, 40% of the mathematics frameworks was not assessed by those instruments. In a slightly different vein, Nevada noted that alignment meant there was consensus among teachers and curriculum specialists that what was being tested was what was intended to be taught.

For most of the states, frameworks and assessments were judged to be aligned if goals and learning objectives were considered in some way in the design or selection of the assessment instruments. Most states lacked a formal and systematic process for determining the alignment among standards, frameworks, and assessments.

METHODS FOR ALIGNING DOCUMENTS

Current practice and the literature suggest three major approaches — sequential development, expert review, and document analysis — for determining whether expectations and assessments are in alignment. These are not the only approaches, nor should they be considered as being applied only in their purest form. In most situations, some combination of the approaches is most applicable.

Sequential Development

Policy elements, such as expectations and assessment, most frequently are aligned by design. A set of standards, for instance, might be converted directly into specifications for developing an assessment. Once one policy element is established, it becomes the blueprint for subsequent elements. The order in which the different documents are prepared can vary, with periodic revision of any one or all documents common. South Carolina had one of the more formal processes of sequential development extending over a period of eight years to reach all of the content areas. For a content area, an in-state committee developed state goals that were revised through public review. Committees of teachers and other educators used these goals to develop drafts of curriculum frameworks and content standards. These drafts were presented for broad public review and input. This type of review helped to assure vertical alignment between the expectations and public opinion. Frameworks and content standards were revised based on the feedback and then went to the State Board of Education for adoption. In fall 1995, the South Carolina State Board of Education approved academic achievement standards for mathematics based on the state's framework. These were measurable outcomes that could be used to develop assessment instruments. Public reviews of the draft science achievement standards were completed in the summer of 1996.

California has had some process for judging the alignment between the state assessment and the curriculum framework for a number of years. Up to around 1983 the same key people who wrote the frameworks, the intellectual and official leadership of the committees, developed the specifications for the assessment including what knowledge of content and what skills would be measured. Booklets entitled *Rationale and Content* were then produced and widely circulated to schools to help them focus on the framework and what would be assessed. After 1983, and with the advent of performance assessment, translating ideas from the framework into test specifications was more difficult. Judgments on the linkage of the assessment with the frameworks became much more dependent on people's judgment (D. Carlson, personal communication, May 22, 1996).

States vary greatly in their policy context. These differences make it difficult to categorize states as using any single model to establish alignment, or in using one model in the same way. In South Carolina, the state legislature had to appropriate funds for the state assessment and, thus, had a strong influence on the form of assessment that would be used by issuing funds only for the form of assessment it approved. The Maryland Board of Education had the responsibility for operating the education system in that state and, with this authority, could institute an assessment system. Maryland's assessment system in 1996 had a strong influence on instructional programs in schools and was considered to be a greater influence on reform than other Maryland policy documents. Georgia had mandated that each district develop its own framework. In July 1994, the Ohio State Board of Education adopted *Ohio's Model Competency-Based Science Program* designed to guide the development of districts' curricula. In states such as Iowa, with less centralized education systems, the state department of education provided technical assistance rather than develop standards or assessments; the authority for establishing alignment in these states resided more at the district and school level than with the state.

Developing standards, frameworks, and assessments in sequence has the advantage of proceeding in a logical process and, after the development of the first document, having known criteria for the development of subsequent documents. Checks of alignment by educators, other experts, and the public can be built into the process. One disadvantage of this approach is the amount of time needed to put a sequentially developed program in place. This approach also ignores a synergism among policy elements: The development of assessments, for example, can provide useful information for thinking about instruction and what students can be expected to learn. For this reason, an iterative process may be more effective than a sequential process. Another disadvantage to sequential development is that it frequently does not reflect reality: In many states, the process for developing expectations and assessments is not linear or sequential, but more dynamic and recursive.

Expert Review

In some states and districts, a panel of experts reviews the policy elements and makes some judgment on their alignment. The formality of the review process will vary. In many states and districts, the development of frameworks and standards is an open process, turning to committees and community forums of teachers, administrators, parents and the public to offer reactions and, in some, to reach consensus. Along with a process of sequential development of documents, as discussed above, most states and districts institute some review process. However, in

sequential development, the more precise content reviews are conducted internally by those who are engaged in the writing or development of the documents. Reviews by external panels are not as frequent. In Michigan, the review for alignment is part of the selection process. Committees are selected to approve and reject items for inclusion in the assessment process based on their fit with the state goals and objectives. The Oregon Department of Education convened a national panel to look at various issues related to its standards (Roeber, 1996). A sub-panel looked at the alignment of the planned assessments and the standards.

Content area specialists are essential to serve as members of any review panel with the purpose of judging the match between expectations and assessment. The number in a review panel can vary, but generally should be five or more. Complex distinctions need to be made requiring a level of sophistication far exceeding general lay knowledge in understanding how students learn. Just judging the quality of either expectations or assessment and their internal consistency can be complex. In a quantitative review of the curriculum frameworks from 12 states, committees of content area specialists found frameworks lacking internal coherence in both style and content, open to misinterpretations from the use of imprecise language, containing poor translations of fundamental concepts and principles from national standards, and inaccurately interpreting learning approaches (Humphrey & Shields, 1996). All of these issues arising from the analysis of one element are confounded in comparing the alignment among multiple elements.

The tools needed for an expert review of documents will vary. A review panel given an opportunity to meet and discuss the match between documents will need less structure by experts than reviewing in separate locations. Remote reviews will require more structure and rating forms to focus reviewers' comments and allow more reasonable aggregation of comments. In either case, reviewers need clear directions of what their task is and what it is not; what will be useful and what will not. Any content analysis or analysis of agreement among complex elements can be very detailed.

A content comparison between a draft of the *National Science Education Standards* (NRC, 1996) and the *Benchmarks for Science Literacy* (AAAS, 1993) illustrate some of the difficulty (AAAS, 1995). The task of describing correspondence between these two documents was time consuming, demanded an extensive familiarity with both documents, and required deciphering two different organizations and aggregations of content. Whereas the *NSES* may provide one statement of a fundamental concept such as the food change, the same ideas were distributed among several different grade levels in the *Benchmarks*.

In another effort, Project 2061 of the American Association for the Advancement of Science experienced the complexity of designing alignment studies. The developers found they needed a minimum of four days to train three-person review teams to produce valid and reliable findings. Teams were trained to use a four-step procedure to analyze curriculum materials and judge their likely contribution to the attainment of specific learning goals as specified in the *Benchmarks* and the *National Science Education Standards*. The four steps included a preliminary inspection, content analysis, instructional analysis, and a summary report. Critical to this analysis were eight distinguishing features — specific learning goals, instruction tied to learning goals, evidence-based arguments, feedback, clarification of learning goals, specific criteria, clarification of criteria, and concrete examples of applying criteria.

Document Analyses

Alignment can be judged by coding and analyzing the documents that convey the expectations and assessments. The Third International (IEA) Mathematics and Science Study (TIMSS) effectively used document analysis to judge the alignment between national and regional curricula and assessment documents (McKnight, Britton, Valverde, & Schmidt, 1992a, 1992b; Schmidt & McKnight, 1995). An analysis of curricula along with the assessment was critical so that educational opportunity could be used as an important link among direct measures of the curriculum, teacher opinions about the content coverage, and achievement data. Initially, repeated ratings two years apart by national committees with the same members produced unstable results in judging the relationship between the curriculum and assessment instruments. The TIMSS staff then adapted a more analytic approach to comparing curricula with assessment.

TIMSS staff devised a three-dimensional grid (content by performance expectations by perspectives) to describe both the curriculum and assessment (Robitaille et al., 1993). They used 10 major content categories for mathematics:

- ✔ numbers;
- ✔ measurement;
- ✔ geometry — position, visualization, and shape;
- ✔ geometry — symmetry, congruence, and similarity;
- ✔ proportionality;
- ✔ functions, relations, and equations;

- ☑ data representation, probability, and statistics;
- ☑ elementary analysis;
- ☑ validation and structure; and
- ☑ other content.

These major categories were divided further into 44 subcategories. Five big categories described performance expectations:

- ☑ knowing,
- ☑ using routine procedures,
- ☑ investigating and problem solving,
- ☑ mathematical reasoning, and
- ☑ communicating.

The five categories delineated the general perspectives:

- ☑ attitudes,
- ☑ careers,
- ☑ participation by under represented groups,
- ☑ interest, and
- ☑ habits of mind.

TIMSS staff used eight broad categories for the science content dimension:

- ☑ earth sciences;
- ☑ life sciences;
- ☑ physical sciences;
- ☑ science, technology, and mathematics;
- ☑ history of science and technology;
- ☑ environmental and resource issues;
- ☑ nature of science; and
- ☑ science and other disciplines.

These categories were divided further into a total of 77 subcategories. The science categories for performance expectations and general perspectives were comparable in number and names used for mathematics, with the addition of safety to the science perspectives.

Trained national committees partitioned the textbooks, curriculum guides, and assessment instruments into blocks defined by changes in content or expected performance by students. The analysis scheme defined a block as a piece of text that was part of a lesson or unit devoted largely to one topic. The different forms of textbook content blocks included a narrative block, graphic block, exercise/question sets, activity blocks, and worked examples. The raters assigned each block a "signature" code determined by the different cells checked on a three-dimensional grid. More than one cell of the grid could be marked for any one block, but usually the number of cells for any one block did not exceed three. The national committees used the same grid to perform a similar analysis on the other document to be compared — an assessment instrument or curriculum guide.

The committees' analyses generated a set of content by performance expectations by perspectives matrices with filled-in cells. Circles (O) marked the block signatures produced by the curriculum guide analysis and plus signs (+) marked the block signatures produced by the assessment analysis. The proportion of cells with both a circle and plus represented the degree of match between the two documents. Cells with only one of the symbols indicated a mismatch. Easily computed measures of reliability determined the quality of the process. In sessions of 22 hours, TIMSS staff successfully trained committees in fifty countries to do this form of document analysis.

One practical issue arose in the document analyses and other data gathering for TIMSS. Teachers, in judging whether assessment activities matched their curriculum, were strongly influenced by the form of the assessment activity. Teachers were more likely to judge an assessment activity to be aligned with the curriculum content demands if the activity had the same form as activities used in the curriculum. They were more likely to reject assessment activities as being aligned with the curriculum if the activities measured the same content, but in an unfamiliar way. Having examples of student work on assessment activities helped to alleviate this problem.

Porter (1995) devised another system for analyzing content to describe enacted curriculum indicators for school improvement. His system used estimates of instructional time spent with students by topic and time spent engaged in cognitive activities to indicate the emphasis given to activities and topics. By estimating the amount of time devoted on an assessment to the same content topics and cognitive categories, the alignment between the assessment instrument and curriculum could be determined. Document analysis requires using a common metric to

compare the curriculum and the assessments. TIMSS used blocks of content and Porter used degree of emphasis based on time.

Quality Control

Each of the three approaches for establishing the alignment among the policy documents lend themselves to different ways of determining their quality. Any of the three approaches can be verified by using one of the other techniques as a check.

Alignment by sequential development is frequently controlled within an agency and will be less likely to have some form of external review. Even though policy elements will undergo public or expert review, the match between the documents may be given less attention or done more informally by staff. Alignment among the documents would be more credible with some external review; however, frequently time lines are so stringent due to legislative mandates or administrative pressures that time is not available for any external review. In the absence of such a review, the assurance of alignment can be strengthened by incorporating checking procedures to be used by agency staff.

The quality of expert reviews will depend on the qualifications and expertise of the reviewers. At least some reviewers need to be very knowledgeable of the content areas and learning. The number of members of a review panel and the mix of the panel will depend on a number of variables — grade ranges, depth of analysis, technical level of the documents, and purposes of the documents. Five members are a minimum to have the important perspectives represented and to obtain some measure of reliability. Providing an opportunity for reviewers to interact with each other and to build consensus will help improve the quality of the review.

Sampling techniques can be used to verify the quality of document analyses. The reliability of partitioning documents into blocks and then coding blocks can be determined by accomplished coders checking a sample of the materials.

Specific Criteria

The ultimate goal for alignment is a fully functional system working towards students learning important mathematics and science. This requires a very deep analysis of both horizontal alignment among the policy elements and vertical alignment with purposes, programs, practices, and student outcomes. Partial progress toward this ultimate goal can be determined by analyzing the alignment among policy elements

using one of the above approaches, some other viable approach, or a combination of these approaches to a desired level of agreement.

Judging alignment is strengthened by using specific criteria to analyze agreement among expectations and assessments. A review of national and state standards and different alignment studies suggested the following categories of criteria — content focus, articulation across grades and ages, equity and fairness, pedagogical implications, and system applicability. A panel of assessment experts, state specialists in curriculum and assessment, and education researchers, convened by CCSSO and NISE, refined further these criteria (Appendix).

The criteria presented here are intended to provide a means for thinking about alignment. They are not considered to be the definitive list. They have not withstood the test of time or use. It is expected that this set of criteria will evolve over time. The criteria are presented in an order first to consider content, then students, then instruction, and finally application to a system. Each criterion is summarized by a brief description of what can be compared between expectations and assessments (unit of comparison) and levels (full, acceptable, and insufficient) for judging the degree of agreement.

CONTENT FOCUS

Expectations and assessments should focus consistently on developing students' knowledge of mathematics and science. This consistency will be present to the extent expectations and assessments share the following attributes:

Categorical Concurrence

The same or consistent categories of content appear in both expectations and assessments. Categorical concurrence is achieved if comparable topic headings and subheadings of content appear in each. This level of detail, however, may vary: standards and frameworks can be statements of general expectations, or they can be more refined description of content. Assessment specifications and activities, designed to provide evidence of students' expected attainment, may be even more specific.

The *National Science Education Standards* (NRC, 1996) used the same eight content topics for each of three grade ranges (K-4, 5-8, and 9-12): Unifying Concepts and Processes, Science as Inquiry, Physical Science, Life Science, Earth and Space Science, Science and Technology, Science in Personal and Social Perspectives, and History and Nature of Science. For an assessment system to have categorical concurrence with these national standards, at least the most general categories of content

assessed should coincide with the eight broad categories of the national standards. Alignment would be even greater if assessment results were reported by those eight categories.

Categorical Concurrence Criteria

UNIT OF COMPARISON

- ✔ *Expectations* Content topics, subtopics, or both, identified by standards or main areas of content specified.
- ✔ *Assessment* Topics by which results are reported (most stringent); subunits topics of instruments; or topics of clusters of assessment activities.

SCALE OF AGREEMENT

- ✔ *Full* A one-to-one correspondence between topics given in expectations and topics by which assessment results are reported.
- ✔ *Acceptable* Assessments cover a sufficient number of topics in expectations so that a student judged to have acceptable knowledge on the assessments will have demonstrated some knowledge on nearly all topics in expectations.
- ✔ *Insufficient* Important topics are excluded from assessments to the extent students can perform acceptably on assessments and still lack understanding of important expectation topics.

Depth of Knowledge Consistency

Depth of knowledge can vary on a number of dimensions, including level of cognitive complexity of information students should be expected to know, how well they should be able to transfer this knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas. The depth of knowledge or the cognitive demands of what students are expected to be able to do is related to the number and strength of the connections within and between mental networks. Understanding in mathematics, for example, is described as making connections between ideas, facts, or procedures (Hiebert & Carpenter, 1992, p. 67). The depth of knowledge required by an expectation or in an assessment is related to the number of connections of concepts and ideas a student needs to make in order to produce a response, the level of reasoning, and the use of other self-monitoring processes. In addition, other factors influence the cognitive demands of performance, including the social or contextual requirements, the variety of representations students are expected to use (written, ver-

bal, pictorial, and variations within each), and requirements for transfer and generalization to new situations.

Expectations and assessments are aligned if what is elicited from students on the assessments is as demanding cognitively as what students are expected to know and do. For example, the *Curriculum and Evaluation Standards for School Mathematics* published by the National Council of Teachers of Mathematics (1989) state that students in grades 9 through 12 should study data analysis and statistics, so that all students can "design a statistical experiment to study a problem, conduct the experiment, and interpret and communicate the outcomes" (p. 167). An assessment system requiring students only to interpret an existing set of data, without designing and experimenting, would fall short and not be aligned with the depth of knowledge specified in this standard. The standard calls on students to identify a problem, draw upon their knowledge of statistics to create a design for an experiment, perform the experiment, organize the information produced, give meaning to this information, and then explain the findings. Assessing students' interpretation of a given data set would require students to demonstrate only a small part of what the standards intend.

Reality needs to influence the comparison of the cognitive demands of expectations as expressed by standards, goals, and objectives with how students are held accountable to perform on assessments for the purpose of alignment. Ideally cognitive studies would be conducted to delineate in some detail what depth of knowledge is required by an expectation and what mental operations students actually used on the corresponding assessments. Such studies can be costly and time consuming. A more realistic analysis would be to seek some expert help and conduct a content analysis using verbs and their objects to judge the match between expectations and assessments. For example, the expectation for a student to "*use and produce a variety of classification systems* to identify organisms" is more demanding cognitively than the expectation to "*discuss the classification* of organisms." Generating a variety of systems generally will require making more connections than discussing what are classifications. The full cognitive demands implied by both statements lack some clarity without more delineation of the specific criteria that will be employed to judge the successful application, use, and production of classification systems or the extent of the discussion of the classification of organisms.

An assessment activity of comparable depth of knowledge to the expectation that students will use and produce a variety of classification systems could be something like this example. A teacher shows her grade

4 students pictures and lists of a range of animals, birds, and insects. Then she asks them to form groups of these by common characteristics and to assign a name and description for each grouping. This assessment activity, in general, would require students to draw upon more specific features of the organisms and require them to sort the organisms in new ways. A satisfactory discussion by elementary students would be a general recognition of the idea of species as a basis for classifying organisms.

Expectations and assessments will be fully aligned if both are cognitively complex or both are cognitively simple. How closely a comparison between expectations and assessments can be made will depend on the specificity of the expectations. The Adams Twelve Five Star Schools, Northglenn, Colorado (1995) clarified a very general statement of standards with more specific indicators of performance. One grade 9-12 science standard stated: "The student knows and is able to demonstrate an understanding that energy appears in different forms, and can move (be transformed), and change (be transformed)" (p. 3).

This statement is too general to really judge what depth of knowledge is sought. The indicators of performance given along with the general statement gave more clarity to what was expected:

- ✔ investigates the quantitative relationships among pressure, volume, and temperatures of gases;
- ✔ characterizes chemical or physical changes as endothermic or exothermic;
- ✔ measures, calculates, and compares voltage, current, and resistance (Ohm's Law);
- ✔ measures, calculates and compares forces and changes in speed (Newton's First and Second Laws);
- ✔ contrasts properties of materials that transmit, reflect, absorb, and/or diffract visible light;
- ✔ examines the effect of lenses, mirrors, diffraction gratings, and opaque objects on visible light; and
- ✔ describes the results of water wave interactions (p. 3).

The verbs used in these indicators of performance — investigate, measure, calculate, compare, contrast — suggest that what is meant by understanding is more than recall of information but requires reasoning, applications of skills, and knowledge of scientific concepts. Assessment aligned with these expectations will have to elicit from students a fairly deep level of knowledge.

Depth of Knowledge Criteria

UNIT OF COMPARISON

- ☑ *Expectations* Rating of most cognitively demanding expected performance for a topic and for all students as determined by number of ideas integrated, depth of reasoning required, knowledge transferred to new situations, multiple forms of representation employed, and mental effort sustained.
- ☑ *Assessment* Rating of most cognitively demanding assessment activity for a topic and taken by all students as determined by number of ideas integrated, depth of reasoning required, knowledge transferred to new situations, multiple forms of representation employed, and mental effort sustained.

SCALE OF AGREEMENT

- ☑ *Full* For each major topic, the most cognitively demanding expected performance for all students is comparable to the most cognitively demanding assessment activity taken by all students.
- ☑ *Acceptable* For nearly all major topics, the most cognitively demanding expected performance for all students is comparable to or can be inferred from the most cognitively demanding assessment activity taken by all students.
- ☑ *Insufficient* Students can be judged as performing at an acceptable level on the assessments without having to demonstrate for any topic the attainment of the most cognitively demanding expected performance for all students.

Range of Knowledge Correspondence

Expectations and assessments cover a comparable span of knowledge within topics and categories. Even with strong categorical concurrence (criterion A), the span of expected knowledge within categories may not always be entirely covered by the assessment system. Tradeoffs frequently have to be made. Time limitations, scoring costs, availability of instruments, and other constraints on assessment can prohibit the measurement of the full range of content or performance expectations for every student. Completely aligned expectations and assessments requires an assessment system designed to measure in some way the full range of expected knowledge within each specified topic. For example, according to the standards published by the Virginia Board of Education (1995), students should be able to read four different types of maps: bathymetric, geologic, topographic and weather. An assessment corresponding to the range of knowledge of this map-reading

expectation would need to measure how well students can interpret information using all four map types.

Thoughtfully designed assessments help to increase the correspondence between the assessed range of knowledge and the expected range of knowledge. Clear assessment specifications can be more helpful in determining that a full range of knowledge is being assessed than the actual assessment instruments. This is particularly true if activities on an instrument represent a selection of possible activities from a conceptual domain defined in detail in the assessment specifications.

One important consideration in judging the correspondence of the range of knowledge is the reference group for reporting results — system, district, school, or individual. Formally assessing the full range of knowledge as described in standards, goals, and objectives for every student in a large system can be extremely difficult, if not impossible. However, with sampling techniques, performance information on the full range of expected knowledge could be gathered for a large group of students. Construct and content domain definitions included in assessment specifications generally identify the range of content, or some decision rule for generating assessment activities, to assure full coverage of content. There will be a high degree of match between expectations and assessments for the total population if the assessment design assures that a reasonable range of knowledge within topics is assessed and appropriate sampling procedures are used so that inferences can be made on this knowledge for the total population. There will be a low degree of match of assessments with expectations for the total population if the designed assessment measures only one or a few special cases of a complex concept or topic.

Alignment between expectations and assessments on the range of knowledge can be achieved in a system even if formal assessments or large-scale assessments do not fully span every topic. Teachers incorporating in their classrooms assessment practices that attend to measuring knowledge within specific topics not fully covered by large-scale assessments contribute towards a fully aligned system.

Range of Knowledge Criteria

UNIT OF COMPARISON

☑ *Expectations* All students are expected to know how to use and all forms of representations they are to be able to use, including major concepts or ideas within and across performance standards, such as type of graphs.

☑ *Assessment* All types or forms of major concepts or ideas included on assessments or in the specifications of the content domains used to select assessment activities.

SCALE OF AGREEMENT

☑ *Full* Students are required on assessments to demonstrate knowledge of all forms or the full range of each major concept or idea expressed in the expected performance.

☑ *Acceptable* Assessment specifications account for nearly all forms or the full range of each major concept or idea expressed in the expected performance so there is a strong likelihood that students' knowledge and use of all forms will be assessed.

☑ *Insufficient* Important forms or specific cases of major concepts and/or ideas given in the expected performance are excluded from or ignored on assessments or their specifications.

Structure of Knowledge Comparability

The underlying conception of science and mathematics knowledge in expectations and assessments are in agreement. For example, if standards indicate that students should understand mathematics "as an integrated whole" (NCTM, 1989) or "science as inquiry" (NRC, 1996), then the assessment activities should be directed toward the same ends. Both expectations and assessments should embody similar requirements for how students are to draw connections among ideas. Assessment of knowledge only as fragmented skills, for example, would not be in full alignment with the national standards.

The preliminary draft of the Illinois Academic Standards (1996) expressed, "Solving problems is at the heart of mathematics. Mathematics is a collection of concepts and skills; it is also a means of investigation, reasoning, and communicating" (p. 25). One of five applications of learning in this document was "Making Academic Connections: Recognize and apply connections of important information and ideas within and among academic learning areas" (p. 26). The depictions of mathematics in these statements imply that students should learn a collection of concepts and skills along with how to solve problems and apply information to other fields. An assessment system aligned with a comparable structure of knowledge as represented by these expectations would have to gather information on students' understanding of specific concepts and skills, ability to solve problems, and application of mathematics to other academic learning areas.

Judging the structure of knowledge comparability, along with many of the other criteria, requires considering the full assessment system. Expectations and assessment are fully aligned according to this criterion if the structure of mathematical understanding expressed in expectations is represented in the assessment system. In the case of judging whether an assessment system is aligned with the Illinois Academic Standards, some part of the system should gather information on how students are able to identify appropriate applications of mathematical ideas both within mathematics and in other fields. For example, students could be asked to develop a mathematical model of a physics experiment. Alignment with the Illinois Academic Standards will be less if little or no part of the assessment system asks students to demonstrate knowledge of how they are able to draw relationships among mathematical ideas and among mathematical ideas with applications to other fields. The standards and assessments will have very low alignment in the structure of knowledge if students are only asked to demonstrate their knowledge of mathematics by recalling or applying isolated concepts, procedures, and skills.

Structure of Knowledge Criteria

UNIT OF COMPARISON

- ☑ *Expectations* Expression of how students are expected to form relationships among ideas such as no relationship (single or isolated ideas), equivalent forms of the same idea, connection of many ideas within the content area, and connection of ideas within the content area and with applications to other areas.
- ☑ *Assessment* Expression of the relationships among ideas students are required to successfully perform on assessments or as indicated on the assessment specifications such as no relationship among ideas, equivalent forms of same idea, connection of many ideas within the content area, and connection of ideas within the content area and with applications to other areas.

SCALE OF AGREEMENT

- ☑ *Full* The relationships among ideas expressed in the performance expectations are the same as those required to perform successfully on the assessments.
- ☑ *Acceptable* Inferences can be made on how each student has formed relationships among ideas as expressed in the performance expectations.

✔ *Insufficient* A student could be judged as fulfilling the expectations without adequate demonstration on assessments that the student is able to draw relationships among ideas as represented by all important structural forms.

Balance of representation

Similar emphasis is given to different content topics, instructional activities, and tasks. The expectations and assessments give comparable emphasis to what students are expected to know, what they should be able to do, and in what contexts they are expected to demonstrate their proficiency. Fulfillment of the four prior criteria requires expectations and assessments to cover comparable topics in the same depth and breadth based on a common understanding of how knowledge is organized. For the balance of representation criterion to be met, the degree of importance of different ideas given in the assessments and expectations should be the same.

Standards frequently describe what all students are expected to know and do without differentiating the priority or emphasis one standard should be given over another. It is assumed, if not stated directly, all of the standards are important. The equal importance of all content standards given in the *National Science Education Standards* is reinforced by statements such as "None of the eight categories of content standards should be eliminated. . . . No standards should be eliminated from a category." (NRC, 1996, pp. 111-112). The New Jersey Mathematics Coalition and the New Jersey Department of Education (1996), in the *New Jersey Mathematics Curriculum Framework*, indicated some differentiation among the 18 standards (Rosentein, Caldwell, & Crown, 1996). This document described four "mathematical processes" standards as the *major* focus for all students — posing and solving problems, communication, making connections, and reasoning. Each of the 10 mathematics content standards was stated as what all students will develop, but without a superlative.

Assessments, as well as curricula, designed to fulfill expectations and standards are constrained by very pragmatic factors such as time, sequencing, and a high variation in the rate of learning. These constraints force those who develop assessments to make decisions about the amount of emphasis or weight that will be given to different topics on a test. A state assessment for grade 5 may be limited to one class period for each content area. If every student has to take the same test during this class period, then tradeoffs will have to be made among the number of open-ended activities, requiring time and extensive effort, and the number of short answer activities, requiring recall or reproduction of information. The psychometric qualities of the test also will need to be considered.

For example, although the national science standards place an equal weight on each of the eight standards for each of three grade ranges, when these expectations are considered in more depth and across grade ranges there is some variation in emphasis. Under the Earth and Space Science category of standards, emphasis for grades K-4 is on developing observation and description skills and basing explanations on observations (NRC, 1996, p. 134). In the middle grades, on the same category, more emphasis is given to constructing models that explain visual and physical relationships (p. 159). For an assessment system to have a comparable balance of representation with the *National Science Education Standards* would require a shift from grades K-4 in which students' knowledge is judged from primarily observing and describing to grades 5-8 in which students' knowledge is judged from primarily drawing inferences from relationships. In the lower grades more emphasis on assessments should be given to determine whether students can observe and describe properties of earth materials and objects in space. In the middle grades, more emphasis on assessments would have to be given to students demonstrating their understanding of the relationship among components of the earth system and using a model of the physical relation among earth, sun, moon, and the solar system to explain such phenomena as the phases of the moon.

Balance of Representation Criteria

UNIT OF COMPARISON

- ✔ *Expectations* Assigned importance on a scale of 100 by topic over full spectrum of performance expectations (total for all topics should be 100).
- ✔ *Assessment* Weight by topic or subtopics for full spectrum of assessments (weight could be determined by the proportion of activities by topic, proportion of average time allocated to do an assessment activity by topic, or according to some other rule).

SCALE OF AGREEMENT

- ✔ *Full* The proportion of assigned importance for topics in performance expectations is equivalent to the weight topics are given on assessments.
- ✔ *Acceptable* Distribution of importance by topics in performance expectations nearly matches the weight of topics in assessments without major exclusions.
- ✔ *Insufficient* Weights on assessment by topic are sufficiently different from the assigned importance in the performance expecta-

tions such that a student could be judged as meeting performance expectations without knowledge of highly emphasized topics.

Dispositional Consonance

When expectations include more than learning concepts, procedures, and their applications — such as molding student attitudes and beliefs about science and mathematics — assessments also should support that broader vision. For example, the *National Science Education Standards* underscore the importance of students becoming self-directed learners. The ability for students to self-assess their understanding is an essential tool for this. Assessment practices aligned with this goal will include opportunities for students to critique their own work and to explain how work samples provide evidence of understanding. To achieve this, teachers need to give students opportunities to reflect on their scientific understanding and abilities, so they can begin to internalize the expectation that they can learn science.

Determination of dispositional consonance between expectations and assessments requires considering the assessments administered on different levels. Classroom assessment of students' habits and attitudes towards science and mathematics frequently are done informally by teachers looking for cues of students' developing dispositions. Even though evidence intimately links affect and cognition (McLeod, 1992), formally measuring individual attitudes in conjunction with achievement is difficult. What may be more easily achieved is obtaining some indication of attitudes and dispositions for large groups. The 1994 North Carolina Competency-Based Curriculum listed as one of five program goals, "Develop responsible attitudes toward the environment, science, technology, and society" (p. 5). The document went on to clarify that "student attitudes **will not** [emphasis in original text] be a part of an achievement [end-of-course or end-of-grade] score" (p. 11). The state assumed the responsibility for measuring how students view science through its state assessments of accumulative knowledge and not at the end of courses. Some degree of alignment with respect to dispositional consonance then, in this case, is achieved by gathering group assessment indicators related to expectations rather than individual assessment indicators.

Dispositional Consonance Criteria

UNIT OF COMPARISON

✔ *Expectations* List of desired dispositions toward the content area students are to develop, including habits, attitudes, and other qualities.

- ✔ *Assessment* List of dispositions toward the content area on which information is gathered, either formally and informally, reported, and used to make decisions about students.

SCALE OF AGREEMENT

- ✔ *Full* Main expected dispositional qualities are observed, monitored, and reported at designated levels within the system.
- ✔ *Acceptable* Some effort is made and attention is given to observing students' development of expected dispositional qualities.
- ✔ *Insufficient* Little or no attention is given to observing or monitoring dispositional qualities even though these qualities are advanced as important; or only easily measured attitudes are incorporated in assessments while more prominently desired qualities are ignored.

ARTICULATION ACROSS GRADES AND AGES

Students' knowledge of mathematics and science grows over time. Expectations and assessments should be rooted in a common view of how students develop, and how best to help them learn at different developmental stages. Over time, students' growing understanding of science concepts and processes will allow them to perform higher levels of analysis and work with a greater tolerance of criticism and uncertainty. Students' increased understanding of number, operations, and generalizations will promote more abstract and algebraic thinking. For expectations and assessments to be aligned, they need to be grounded in a similar view of cognitive development and advancement in knowledge. This common view should be based on:

Cognitive Soundness Determined by Best Research and Understanding There has been considerable research on the learning of mathematics and science, which has produced extensive knowledge of how students mature in their understanding of these content areas (Romberg & Carpenter, 1986; Stein, Grover, & Henningsen, 1996). Expectations and assessments should build on this knowledge to develop a sound learning program, and they should do so in ways that are aligned. Students' knowledge of mathematics can be thought of as internal networks of representations (Hiebert & Carpenter, 1992). Understanding is built gradually as new information is connected to existing networks of ideas. According to this view of understanding, new networks are constructed by reconfiguring old networks and forming new connections. Thinking and reasoning develop along with understanding of routine skills

(Resnick & Resnick, 1992). The alignment between expectations and assessments will be strengthened if they both are grounded in a common view of how understanding is developed, such as through continual expansion of networks of representations.

The preliminary draft of the Illinois Academic Standards (1996) for mathematics listed for each academic standard learning benchmarks to further delineate across five grade ranges what was expected for students. Benchmarks for a measurement standard — measure and compare quantities using appropriate units, instruments and methods — represented a developing understanding of measurement ideas over four grade ranges. Early elementary students were to measure quantities with customary and metric systems, relating a metric to a quantity. Late elementary students were to compare and convert units of measures of quantities, expanding their internal "measurement networks" to relationships among different forms of units. Middle or junior high school students were to apply the concepts and attributes of measures to practical situations, forming new links between their understanding of measurement to useful applications. Early high school students were to apply units, domains or ranges and scales to describe and compare functions, numerical data and physical objects, expanding the "measurement networks" even further to more abstract quantities. For an assessment system to be aligned with this measurement standard, assessment instruments used over these grade levels need to be based on a similar view of how students grow in understanding of measurement from forming single relationships between units and quantities to a more complex network of ideas incorporating more abstract units and quantities.

Cognitive Soundness Criteria

UNIT OF COMPARISON

- ✔ *Expectations* The expressed or implied underlying understandings and theory of how students' learning matures over time (behavioral, constructive, developmental levels, combination, etc.) and accepted individual differences among students.
- ✔ *Assessment* Analysis across grades of the implied theory (theories) of learning represented in assessment system instruments, procedures, and attention to individual differences.

SCALE OF AGREEMENT

- ✔ *Full* Assessment instruments for each grade level are developmentally appropriate and, across levels, represent a reasonable progression in learning as depicted or implied in the expectations.

- ✔ *Acceptable* Assessment instruments for each grade level represent most, but not all, developmental expectations and generally follow the same progression in learning depicted or implied in the expectations.

- ✔ *Insufficient* Assessment instruments include activities out of sync with the explicit or implied developmental expectations and the sequence of instruments across grades does not depict the same progression of learning as stated or implied in the expectations so that students' knowledge is overstated or understated.

Cumulative Growth in Content Knowledge During Students' Schooling
Expectations and assessments should be linked by underlying rationales of mathematics and science as content areas. Although the learning of mathematical and scientific concepts over time doesn't follow a strict order of steps, students often need to grasp certain concepts and ideas in order to address more advanced ideas. In order for high school students to take part in scientific inquiry, for example, they first need to learn to identify questions and concepts that guide scientific investigations, to design and conduct such investigations, to use technology and mathematics, to formulate and revise scientific explanations, and to recognize and evaluate alternative explanations. For high school students to incorporate more abstract knowledge, such as the structure and function of DNA, they need to have developed in grades K-8 foundational understanding of life sciences. Aligned expectations and assessments describe and represent, in complementary fashion, the underlying structure of content knowledge students need to develop, and how their instructional experiences should be organized. Assessments at a given grade level should reflect whether students are expected by that grade level to have a foundational knowledge, expanding knowledge, or fully mature knowledge of the assessed concept or idea.

Cumulative Growth in Content Knowledge Criteria

UNIT OF COMPARISON

- ✔ *Expectations* The expressed or implied understanding of how students' knowledge of content will be structured and will mature over time.

- ✔ *Assessment* Analysis across grades of how the growth in understanding of content ideas and the relationships among content ideas are represented in assessment system instruments and procedures.

SCALE OF AGREEMENT

- ✔ *Full* Assessment instruments elicit information compatible with how students' knowledge of content ideas develops over time and how students relate these ideas with each other as reflected in expectations.
- ✔ *Acceptable* Assessment instruments elicit information according to general patterns of how students' knowledge of content ideas develops over time and how students relate these ideas with each other as reflected in expectations.
- ✔ *Insufficient* Assessment instruments across the grades do not represent a logical or sequential growth in student content knowledge over time implied in the expectations. Assessments in lower grades require a more advanced understanding of ideas than do those in later grades as depicted in the expectations. Or, important stages or indicators in the development of content areas as depicted in the expectations are excluded from the assessment system.

EQUITY AND FAIRNESS

When expectations are that all students can learn to high standards, aligned assessments must give every student a reasonable opportunity to demonstrate attainment of what is expected. Expectations and assessments that are aligned will serve the full diversity in the education system through demanding equally high learning standards for all students while fairly providing means for students to demonstrate the expected level of learning. The knowledge a student will demonstrate on an assessment can vary by the form of assessment (Baxter, Shavelson, Herman, Brown, & Valadez, 1993). Even a slight variation in the context of a question can alter performance, such as giving a bare arithmetic problem or the same mathematical content in a word problem (Clements, 1980; Van den Heuvel-Panhuizen, 1996). Rarely will one form of assessment be capable of producing valid evidence for all students. For example, only half of students who answered a multiple-choice question on percents given on the 1986 National Assessment of Education Progress gave a satisfactory response to a following open-ended question that asked them to explain their response (Gay & Thomas, 1993). Students judged to have understood the concept by a correct response to the multiple-choice question had a clear misunderstanding of the concept as noted by their written response.

A student's ability to perform well on a particular assessment can depend on a number of factors in addition to the level of knowledge, including

culture, social background, and experiences (Santos, Driscoll, & Briars, 1993). Therefore, expectations and assessments will be better aligned, and more equitable, if alternative forms of assessment to measure student attainment and procedures are in place to assure that the knowledge of each student has been fairly assessed. The challenge becomes developing and maintaining an aligned system with a variety of means of assessment that ensures each student's full attainment of an expectation is measured. Students' unsuccessful demonstration on one measure could be an interaction with the form of assessment rather than the content. An aligned assessment system will provide opportunities for students who are unsuccessful on one form to demonstrate their understanding by taking an alternative form or taking the same form under different conditions as long as validity and reliability are maintained.

Assessment of expectations that are more open-ended and require students to apply knowledge to their life experiences will produce very diverse responses. Grades 3-6 students in South Carolina were expected to "participate in problem-solving activities through group and individual investigations so that they can relate the use and understanding of numeration systems to their world" (South Carolina State Department of Education, 1993, p. 48). Because students' life experiences are very different, their demonstration of the application of numbers to their world can vary greatly. For assessment practices to be equitably and fairly aligned with expectations, such as this one from South Carolina, student responses need to be judged on the adequacy of the application of numbers and not on what experiences students portray, some of which may not be within the experiences of the person scoring the assessment. The South Carolina Mathematics Framework emphasized this point under its principles and goals for mathematics assessment, "To be fair to all, assessments must be sensitive to cultural, racial, and gender differences" (p. 110). Expectations for students to apply their knowledge of mathematics and sciences can be met in many different ways. For assessments to be aligned with these expectations, they also need to be similarly robust so as not to misrepresent what science and mathematics a student truly knows or can do. An underlying principle is to maximize the participation of students achieving the expectations and demonstrating on assessments what is the full extent of their knowledge.

It may be difficult to gauge alignment between expectations and assessments on the criteria of fairness and equity until both have been in place for some time. Consistently low scores over time on an assessment of a particular learning goal may be the result of many factors, including misplaced expectations, rather than poor instruction or lack of effort by students. Students may be developmentally unprepared to attain a par-

ticular expectation, for example, or the structure of the curriculum may keep them from attaining sufficient experiences to learn what is expected. Time is required for patterns to form in order to decipher how expectations and assessments are working in concert with each other to be equitable and fair.

Equity and Fairness Criteria

UNIT OF COMPARISON

- ✔ *Expectations* Levels of knowledge that all or different groups of students are expected to achieve by specific times (e.g., content knowledge understanding by all students by end of grade 12, by college-intending students by grade 12 . . .).
- ✔ *Assessment* Degree to which the assessment system affords students a reasonable opportunity to demonstrate the full level of knowledge as expected for all students or for their specific group.

SCALE OF AGREEMENT

- ✔ *Full* Students are afforded a fair and reasonable opportunity to demonstrate the full level of knowledge expected for all students. Assessment practices are such that variation of assessment results are only a variation in the attainment of expectations and free from being influenced by culture, ethnicity, gender, or any other irrelevant factor.
- ✔ *Acceptable* Assessment practices are appropriate to measure the attainment of expectations for all or any designated group of students while minimizing culture, ethnicity, or gender bias.
- ✔ *Insufficient* Important judgments are made on students' attainment of expectations based on biased or limited assessment practices that do not afford a student or group of students a reasonable opportunity to demonstrate what they know and can do.

PEDAGOGICAL IMPLICATIONS

Classroom practice greatly influences what students learn. Expectations and assessments can and should have a strong impact on these practices, and should send clear and consistent messages to teachers about appropriate pedagogy.

Judging the pedagogical implications of expectations and assessments requires more than simple content analysis. Any such review must attempt to gauge the likely implications on pedagogy. Meaningful analy-

ses have been done by directly asking teachers how they interpret expectations and assessments, and how their classroom practices fit with them (Cohen, 1990; Romberg, Zarinnia, & Williams, 1990).

Of course, the true test is what happens in the classroom. For example, educators are now paying increased attention to the importance of involving students in scientific inquiry, hands-on learning, and more "authentic" instruction (Newmann, Secada, & Wehlage, 1995). Assessments that suggest a more passive type of instruction would be less aligned with those expectations. Likewise, expectations that indicate students are to develop abilities to perform scientific inquiry through actively constructing ideas and explanations (NRC, 1996) will lack full alignment with assessments that are solely based on an assumption that students have memorized the canonical ideas and explanations. Alignment is achieved when the instructional practices and materials implied by expectations and those implied by assessments are consistent.

Critical elements to be considered in judging alignment include:

Engagement of Students and Effective Classroom Practices
Traditional forms of student assessment and the constraints imposed by limits on time and other resources may place an inordinate influence on the superficial acquisition of skills and facts or any one area in relationship to others. In this way, education systems can gravitate toward readily measured outcomes, instead of more complex but also more desirable outcomes, such as students being able to investigate, create models, or otherwise demonstrate deeper content knowledge (McCarthy, 1994).

Expectations and assessments need to work together to provide consistent messages to teachers, administrators, and others about the goals of learning activities. For example: A preliminary draft of statewide academic standards for Illinois indicated that students should learn and contribute productively both as individuals and as members of groups (Illinois Academic Standards Project, 1996). This was defined in the draft as an important skill that will greatly determine the success of students later in life. But if no part of the assessment system produces evidence of whether students are contributing productively as members of groups, then teachers would receive conflicting messages about how much classroom time should be spent having students work in teams.

Engagement of Students and Effective Classroom Practices Criteria

UNIT OF COMPARISON

✔ *Expectations* Range of instructional practices most likely for all students to achieve the full extent of expectations.

- ✔ *Assessment* Range of instructional practices most likely for all students to develop the knowledge and experiences to perform satisfactorily on assessments.

SCALE OF AGREEMENT

- ✔ *Full* Instructional practices most likely to have students fully achieve expectations are the same as the instructional practices most likely to have students adequately demonstrate their attainment of these expectations on available assessments.
- ✔ *Acceptable* Students are not disadvantaged by instructional practices required for them to do well on assessments in achieving the full extent of the expectations.
- ✔ *Insufficient* In order for students to do well on assessments, teachers must give undue emphases to instructional practices that inhibit students' learning content to the full extent as expressed by the expectations.

Use of Technology, Materials, and Tools

Technology, materials and tools are vital to knowing and "doing" mathematics and science today. Students should develop skill and confidence using tools such as calculators and computers in their everyday lives (National Council of Teachers of Mathematics, 1991). In science they should increase their repertoire of tools and techniques and improve their skills in measurement, calculation, and communications (American Association for the Advancement of Science, 1993).

The draft of the *Texas Essential Knowledge and Skills* for science (Texas Education Agency, 1996) stratified expectations as basic understanding, knowledge and skills, and performance descriptions. One basic understanding for middle grades was that change occurs and can be observed and measured. A knowledge and skills goal for this basic understanding was for students to know that theoretical work done by early scientists has revolutionized modern science. This goal was delineated further by performance descriptions including:

The sixth-grade student, using resources·including books, periodicals, videos, technology, and experts, identifies the contributions of different scientists from different cultures at different times (p. 27).

Fully assessing students' attainment of this performance expectation requires determining whether students can use these resources to learn what different scientists from different cultures have done. Lessening the requirement on an assessment by having students respond to ques-

tions rather than actually having them use materials could result in lowering the priority for students to use multiple resources in doing research. Such an assessment would not be aligned with the expectation and could influence teachers to give less attention to students meeting the full intent of the expected performance.

Expectations and assessments need to send students consistent messages about technology and how it is related to what they are expected to learn. If standards indicate that students should learn to routinely use calculators or computers, for example, then the curriculum should provide adequate opportunity for students to use them in this manner. To be aligned, assessments should allow students to use calculators and computers effectively to derive correct answers.

Use of Technology, Materials, and Tools Criteria

UNIT OF COMPARISON

- ☑ *Expectations* The technology, materials, and tools students need to be skilled in using to achieve the full extent of expectations.
- ☑ *Assessment* The technology, materials, and tools students need to be skilled in using to perform satisfactorily on assessments.

SCALE OF AGREEMENT

- ☑ *Full* Adequate performance on assessments requires students to be accomplished in using the full range of technology, materials, and tools as intended by the expectations.
- ☑ *Acceptable* Adequate performance on assessments requires students to be accomplished in using a sampling of technology, materials, and tools as intended by the expectations.
- ☑ *Insufficient* Students are prohibited on assessments from using technology, materials, or tools that students are expected to become accomplished in using.

SYSTEM APPLICABILITY

Although expectations and assessments should seek to encourage high expectations for student performance, they also need to form the basis for a program that is realistic and manageable in the real world. The policy elements must be in a form that can be used by teachers and administrators in a day-to-day setting. Also, the public must feel that these elements are credible, and that they are aimed at getting students to learn important and useful mathematics and science. Expectations and assess-

ments are in alignment based on system applicability if those who have a stake in the education system, those who are to implement the system, and those who are to be held accountable are able to understand these documents, see how they are related, and believe they are attainable. Public review and discussion of expectations has been used by many states to gain public buy-in of expectations. Having assessment as an open process (NCTM, 1995) can further public understanding and acceptance of how attainment of expectations will be measured. Ongoing review of how the expectations and assessments are working as a system and as intended will help assure system applicability.

System Applicability Criteria

UNIT OF COMPARISON

- ✔ *Expectations* Degree all important system stakeholders understand, accept, and value expectations; and the degree they think expectations are attainable.
- ✔ *Assessment* Degree all important system stakeholders understand, accept, and value assessments; and the degree they think assessments measure important knowledge and skills.

SCALE OF AGREEMENT

- ✔ *Full* The public, teachers, students, and others within the system view expectations and assessments as closely linked, acceptable, attainable, and important.
- ✔ *Acceptable* The public, teachers, students, and others within the system do not fully understand the link between expectations and assessments, but are favorable towards them and are willing to support students' attainment of them.
- ✔ *Insufficient* The public, teachers, students, and others see little relationship between expectations and assessments and give weight to one over the other.

Conclusions

These five categories are intended to be a comprehensive set for judging the alignment between expectations and assessments. Each general category and all subcategories are important in ascertaining the coherence of a system — the degree that assessments and expectations converge to direct and measure student learning. In practice, to reach full agreement between expectations and assessments on all criteria is

extremely difficult. Tradeoffs will need to be made because real constraints exist on any education system such as resources, finances, time, and legal authority. Decisions on what tradeoffs are to be made among these criteria or on what level of compliance will be acceptable should be made in full awareness of potential consequences.

How decisions on tradeoffs and lessening compliance will be made will depend on a number of factors. The assessment of content knowledge in depth sometimes prohibits assessment of the full range of content. Sampling of content is generally required on most forms of assessment. The burden falls on those in the system to justify what is acceptable and reasonable within the existing context. In any case, assessment should be thought of very broadly including system-wide, local, and classroom assessments. The assessment of the attainment of some expectations to the depth and breadth required most likely will be the responsibility of classroom teachers. The underlying principle is to have a high degree of match between what students are expected to know and what information is gathered on students' knowledge. If the criteria are not fully met, this should be done in full awareness of what action is being taken and to what degree alignment is weakened.

Above all else, when judging the alignment within a system and using these criteria to consider the relationship of expectations and assessments, a sense of reality needs to be maintained. The available resources, the amount of time required, legislative mandates, and other factors will influence how well alignment can be determined, and how practical it is to make such determinations.

Alignment of expectations and assessments is a key underlying principle of systemic and standards-based reform. Establishing alignment among policy elements is an early activity for improving the potential for realizing significant reform. Those working to build aligned systems should not think too narrowly about the task. The criteria presented here demonstrate that a number of factors can be considered in judging alignment among policy elements. These can be studied in several alternative and potentially complementary ways.

In approaching reform, the consideration of alignment cannot come too soon. And just as educators need to remain vigilant to assure that expectations, assessments, and instructional practices are current, they also will need to review the alignment among these major policy elements as new policies are instituted, new administrative rules are imposed, and system needs are changed.

References

Adams Twelve Five Star Schools. (1995). *Science curriculum framework grades 9-12*. Northglenn, CO: Author.

American Association for the Advancement of Science (Project 2061). (1993). *Benchmarks for science literacy: Science for all Americans*. New York: Oxford University Press.

American Association for the Advancement of Science. (1995). *Summary comparison of content between the November draft of National Science Education Standards and Benchmarks for science literacy: Science for all Americans*. Washington, DC: Author.

Baker, E. L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 131-153). Englewood Cliffs, NJ: Prentice Hall.

Baxter, G. P., Shavelson, R. J., Herman, S. J., Brown, K. A., & Valadez, J. R. (1993). Mathematics performance assessment: Technical quality and diverse student impact. *Journal for Research in Mathematics Education, 24*(3), 190-216.

Blank, R. K., & Pechman, E. M. (1995). *State curriculum frameworks in mathematics and science: How are they changing across the states?* Washington, DC: Council of Chief State School Officers.

Bybee, R. W. (1995). Achieving scientific literacy. *The Science Teacher, 62*(10): 28-33.

Chubin, D. E. (in press). Systemic evaluation and evidence of education reform. In D. Bartels & J. O. Sandler (Eds.), *Implementing science education reform: Are we making an impact?* Washington, DC: American Association for the Advancement of Science.

Clements, M. A. (1980). Analyzing children's errors on written mathematical tasks. *Educational Studies in Mathematics, 11*, 1-21.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*(3), 327-345.

Consortium for Policy Research in Education. (1991). *Putting the pieces together: Systemic school reform* (CPRE Policy Briefs). New Brunswick: Rutgers, The State University of New Jersey, Eagleton Institute of Politics.

Council of Chief State School Officers. (1996). *Key state education policies on K-12 education: Content standards, graduation, teacher licenses, time and attendance*. Washington, DC: Author.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Ferrini-Mundy, J., & Johnson, L. (1996). Building the case for standards-based reform in mathematics education. In S. Bartels (Ed.), *Implementing science education reform: Are we making an impact? AAAS Yearbook (36)*. Washington, DC: American Association for the Advancement of Science.

Gay, S., & Thomas, M. (1993). Just because they got it right, does it mean they know it? In N. L. Webb & A. F. Coxford (Eds.), *Assessment in the mathematics classroom. 1993 Yearbook* (pp. 130-134). Reston, VA: National Council of Teachers of Mathematics.

Harmon, M. (1991). Fairness in testing: Are science education assessments biased? In G. Kulm & S. M. Malcom (Eds.), *Science assessment in the service of reform* (pp. 31-54). Washington, DC: American Association for the Advancement of Science.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 65-97). New York: Macmillan.

Humphrey, D. C., & Shields, P. M. (1996). *A review of mathematics and science curriculum frameworks*. Menlo Park, CA: SRI International.

Illinois Academic Standards Project. (1996). *Preliminary draft: Illinois academic standards for public review and comment, English language arts and mathematics, Volume One, State goals 1-10*. Springfield, IL: Author.

Laguarda, K. G., Breckenridge, J. S., Hightower, M. M., & Adelman, N. E. (1994). *Assessment programs in the Statewide Systemic Initiatives (SSI) states: Using student achievement data to evaluate the SSI*. Washington, DC: Policy Studies Associates.

Madaus, G. F. (1983). *The courts, validity, and minimum competency testing*. Boston: Kluwer-Nijhoff.

Mathematical Sciences Education Board. (1993). *Measuring what counts. A conceptual guide for mathematics assessment*. Washington, DC: National Academy Press.

McCarthy, C. (1994). Being there — A mathematics collaborative and the challenge of teaching mathematics in the urban classroom. In N. L. Webb & T. A. Romberg (Eds.), *Reforming mathematics education in America's cities. The Urban Mathematics Collaborative Project* (pp. 173-195). New York: Teachers College Press.

McKnight, C., Britton, E. D., Valverde, G. A., & Schmidt, W. H. (1992a). *Survey of mathematics and science opportunities: Research report series No. 42: Document analysis manual*. East Lansing: Michigan State University, Third International Mathematics and Science Study.

McKnight, C., Britton, E. D., Valverde, G. A., & Schmidt, W. H. (1992b). *Survey of mathematics and science opportunities: Research report series No. 43: In-depth topic trace mapping, draft 4*. East Lansing: Michigan State University, Third International Mathematics and Science Study.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 575-596). New York: Macmillan.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13- 103). New York: Macmillan.

Messick, S. (1994, March). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258.

National Academy of Sciences. (1997). *Preparing for the 21st century. The education imperative*. Washington, DC: Author. (http://www2.nas.edu/21st)

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

Newmann, F. M. (1993). Beyond common sense in educational restructuring: The issues of content and linkage. *Educational Researcher*, 22(2), 4-13, 22.

Newmann, F. M., Secada, W. G., & Wehlage, G. G. (1995). *A guide to authentic instruction and assessment: Vision, standards, and scoring*. Madison, WI: Center on Organization and Restructuring of Schools.

Office of Technology Assessment. (1988). *Technology and the American transition*. Washington, DC: U.S. Government Printing Office.

Porter, A. C. (1995). *Developing opportunity-to-learn indicators of the content of instruction: Progress report*. Madison, WI: Wisconsin Center for Education Research.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of attitude, achievement and instruction* (pp. 37-75). Norwell, MA: Kluwer.

Robitaille, D., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E., & Nicol, C. (1993). *The third international mathematics and science study: Monograph No. 1: Curriculum frameworks for mathematics and science*. Vancouver, Canada: Pacific Educational Press.

Roeber, E. D. (1996). *Review of the Oregon content and performance standards. A report of the National Standards Review Team prepared for the Oregon Department of Education*. Salem: Oregon Department of Education.

Romberg, T. A., & Carpenter, T. P. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 850-873). New York: Macmillan.

Romberg, T. A., Zarinnia, E. A., & Williams, S. (1990). *Mandated school mathematics testing in the United States: A survey of state mathematics supervisors*. Madison, WI: National Center for Research in Mathematical Sciences Education.

Rosenstein, J. G., Caldwell, J. H., & Crown, W. D. (1996). *New Jersey Mathematics Curriculum Framework: A collaborative effort of the New Jersey Mathematics Coalition and the New Jersey Department of Education*. New Brunswick, NJ: The New Jersey Mathematics Coalition, Rutgers, The State University of New Jersey.

Santos, M., Driscoll, M., & Briars, D. (1993). The classroom assessment in a mathematics network. In N. L. Webb & A. F. Coxford (Eds.), *Assessment in the mathematics classroom. 1993 Yearbook* (pp. 220-228). Reston, VA: National Council of Teachers of Mathematics.

Schmidt, W. H., & McKnight, C. (1995). Surveying educational opportunity in mathematics and science: An international perspective. *Educational Evaluation and Policy Analysis, 17*(3), 337-353.

South Carolina State Department of Education. (1993). *South Carolina curriculum frameworks*. Columbia, SC: Author.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488.

Texas Education Agency. (1996). *Science: Texas essential knowledge and skills*. Austin, TX: Author.

U.S. Congress, House. (1994, September 28). Improving America's Schools Act. Conference report to accompany H. R. 6 Report 103-761. Washington, DC: U.S. Government Printing Office.

Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Utrecht, The Netherlands: Freudenthal Institute.

Virginia Board of Education. (1995). *Standards of learning for Virginia public schools*. Richmond, VA: Author.

Webb, N. L. (1993). Mathematics education reform in California. In *OECD Documents: Proceedings of a conference. Science and Mathematics Education in the United States: Eight innovations* (pp. 117-142). France: OECD.

Zucker, A. A., Shields, P. M., Adelman, N., & Powell, J. (1995). *Evaluation of NSF's Statewide Systemic Initiatives (SSI) program: Second year report, Cross-cutting themes*. Menlo Park, CA: SRI International.

Glossary

Assessment Activity — a specific sample of questions that will elicit a sample of student behavior.

Assessment Framework — a list of expectations that will be assessed.

Assessment Instrument — a purposeful collection of assessment activities intended to measure one or more concepts or a range of knowledge.

Assessment Report — the summary (numerical or qualitative) of student performance on an assessment instrument. May be reported at the student, classroom, school, district, state, or other levels.

Assessment Specifications — specific aspects, limits, and boundary conditions on the domain of knowledge being assessed to guide the selection and development of assessment activities.

Assessment Standards — a series of criteria for judging the adequacy of an assessment of student achievement.

Benchmarks — more specific levels of desired performance at various grade levels or ranges of grades. Benchmarks also is used as another term for performance standards, or more rarely, content standards.

Blueprint — an overall description of the assessment to be created. This will include what will be assessed, how the assessment will be created and validated, anticipated reports of results, and anticipated uses of results. The document may also have sample assessment exercises and sample scoring rubrics.

Curriculum Framework — a document that specifies what students should know and be able to do for grade levels or grade ranges and the organization of the curriculum and instruction.

Performance Standard — defines levels of quality of student performance on an overall assessment instrument. Typically, two or three levels of desired overall performance are defined.

Scoring Rubric — a set of rules for assigning different levels of performance, ordered by quality, to student responses on open-ended assessment activities.

Appendix: Task Force Participants: Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education

Andrew Porter, Task Force chair, Director, Wisconsin Center for Education Research

Rolf Blank, Director of Education Indicators, Council of Chief State School Officers

Joyce Krumtinger, Mathematics Education, Illinois Department of Education

Mozell Lang, Science Education, Michigan Department of Education

Donna Long, Mathematics Education, Indiana Department of Education

Megan Martin, Science Education/Assessment Consultant, Council of Chief State School Officers

Senta Raizen, Director, National Center for Improving Science Education

Doris Redfield, Student Assessment, Virginia Department of Education

Ed Reidy, Deputy Commissioner, Kentucky Department of Education

Ed Roeber, Student Assessment, Council of Chief State School Officers

Eleanor Sanford, Student Assessment, North Carolina Department of Education

Walter Secada, Mathematics Education, Wisconsin Center for Education Research

Sharif Shakrani, Assessment Division, National Center for Education Statistics

Linda Sinclair, Science Education, South Carolina Department of Education

William Tate, Mathematics Education, Wisconsin Center for Education Research

Roger Trent, Student Assessment, Ohio Department of Education Research

Norman Webb, Mathematics Education/Assessment, Wisconsin Center for Education Research

David Wiley, Education and Social Policy, Northwestern University

EX OFFICIO

Larry Suter, Program Officer, National Science Foundation

IMPLEMENTING THE ADEQUATE YEARLY PROGRESS PROVISIONS OF TITLE I IN THE IMPROVING AMERICA'S SCHOOLS ACT OF 1994

by Phoebe Winter

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
WASHINGTON, DC

State Collaborative on Assessment and Student Standards Comprehensive Assessment Systems for IASA Title I/Goals 2000 Study Group on Adequate Yearly Progress

ADRIENNE Y. BAILEY
Council of the Great City Schools

EVERETT BARNES
RMC Research Corporation

DALE CARLSON
California Department of Education

RUBEN CARRIEDO
San Diego Public Schools

JANET CARROLL
Rhode Island Department of Education

JONATHAN DINGS
Kentucky Department of Education

LINDA HANSCHE
Georgia State University

REBECCA KOPRIVA
Delaware Department of Public Instruction

ROBERT LINN
University of Colorado

LINDA LOWNES
Los Angeles Unified School District

CAROLE PERLMAN
Chicago Public Schools

PHYLLIS MCCLURE
Title I Reform Network

EDWARD ROEBER
Council of Chief State School Officers

GRACE ROSS
U.S. Department of Education

JEAN SLATTERY
Rochester City School District

CONNIE WESSON
Los Angeles Unified School District

DAVID WILEY
Northwestern University

PHOEBE WINTER
Council of Chief State School Officers

The assistance of Cindy Brown and Julia Lara of CCSSO's Resource Center on Educational Equity is gratefully acknowledged.

© Copyright 1996 by the Council of Chief State School Officers. This document may not be reproduced without the permission of the Council of Chief State School Officers, the copyright holder. The views expressed in this publication are those of the author and do not necessarily represent those of the Council of Chief State School Officers or the SCASS project states.

This report is part of a series sponsored by the Council of Chief State School Officers' State Collaborative on Assessment and Student Standards (SCASS), Comprehensive Assessment Systems for IASA Title I/Goals 2000. The series addresses issues related to the standards and assessment provisions of Title I.

Two reports in the series focus on requirements related to determining the yearly progress of schools and districts. This paper is designed to highlight some of the issues explored in depth in the monograph, *Adequate Yearly Progress in Title I of the Improving America's Schools Act of 1994: Issues and Strategies,* written by Dale Carlson. The longer paper provides detail on how to use a standards-based approach to measure school and local education agency (LEA) progress in serving children covered by Title I. Both papers were written with the advice of and input from the Study Group on Adequate Yearly Progress, composed of state and local educational agency staff, CCSSO staff, university faculty, a representative from the Title I Reform Network, and a representative from the Council of the Great City Schools.

Overview

Title I of the Improving America's Schools Act of 1994 requires states to develop or adopt challenging content and student performance standards in at least mathematics and reading/language arts and measure school and district progress in enabling students served by Title I to meet these standards. Performance standards must describe at least two levels of high performance, advanced and proficient, and a third level, partially proficient. States must develop or adopt annual assessments to measure student performance on the mathematics and reading/language arts standards. The results of these assessments are to be the major source of information in determining whether schools and districts are making adequate progress in educating students served by Title I. Student content standards and performance standards must be in place by the 1997-1998 school year, and assessments based on these standards must be in place by the 2000-2001 school year. In the interim, states are required to devise and implement procedures for identifying schools in need of improvement, based on accurate information about yearly academic progress.

If a state has or will develop/adopt content standards and student performance standards that apply to all students, then these same standards must apply to students served by Title I. Assessments that the state uses

to measure all students must be used to measure the achievement of students served by Title I. These provisions were put in place so that states would have a single system of standards and assessments for all students, rather than requiring the development of supplementary systems solely for the purpose of meeting requirements of Title I. These provisions also require that students served by Title I be educated to meet the same standards expected of all children, rather than holding these students to separate (and usually lower) standards and expectations.

IASA statutes and final regulations do not explicitly mention the possibility of states' allowing the use of locally developed content standards, performance standards, or assessments for Title I purposes, in lieu of a state-developed system. However, because IASA is linked to Goals 2000, the local flexibility options in Goals 2000 apply to Title I of IASA. States may choose to permit LEAs to create standards and assessments that meet or exceed state models or criteria and use these standards and/or assessments to meet Title I requirements. This flexibility allows states to use one of three approaches: a state-level approach, an approach that combines local and state-level standards and assessments, or an approach that relies solely on local standards and assessments, with state models or criteria to guide their development.

The state has the responsibility for determining which of the three approaches to standards and assessments it will use. If the state chooses to allow LEAs to use local standards and/or assessments, the state is responsible for ensuring that these standards and assessments are at least as rigorous as the state's, if there are state-level standards and assessments. If there are no state-level standards and assessments, the state must provide models or criteria for ensuring that local standards and assessments are rigorous and that they will apply to all students. States permitting this local flexibility, either through allowing the use of local standards and assessments or through combining state and local information, must meet several conditions. States must ensure that local standards are rigorous, meet or exceed state standards, and are assessed in a manner comparable across the state.

Assessments used to measure school and LEA progress for Title I purposes must

- ✔ be administered yearly at some time during grades 3 through 5, grades 6 through 9, and grades 10 through 12;
- ✔ include multiple up-to-date measures;
- ✔ include the measurement of higher order thinking skills;

- ✔ provide adaptations and accommodations for children with diverse learning needs; and
- ✔ include limited English proficient students, and assess in the language most likely to yield accurate and reliable information in subjects other than English.

The assessments must yield scores or other information on attainment of content standards at the student level. Results must be able to be disaggregated within each state, LEA, and school and be reported (if "statistically sound," at the school and LEA levels[1]) by gender, major racial and ethnic groups, English proficiency status, migrant status, students with disabilities as compared to students without disabilities, and economically disadvantaged students as compared to students who are not economically disadvantaged.

Adequate Yearly Progress

Title I emphasizes the responsibility of schools and LEAs receiving Title I funds in educating their students to meet the state standards. Schools and LEAs are accountable for the amount of improvement they make in each year in educating children served by Title I, as reflected in the statute (Section 1111(b)(2)(B)):

Adequate yearly progress shall be defined in a manner —

(i) that is consistent with guidelines established by the Secretary that result in continuous and substantial yearly improvement of each local educational agency and school sufficient to achieve the goal of all children served under this part meeting the State's proficient and advanced levels of performance, particularly economically disadvantaged and limited English proficient children.

Like the requirements for standards and assessments, the requirements for measuring the progress of schools and LEAs are designed to allow states to collect information needed for Title I within a single system of accountability. A state may measure progress for Title I purposes within the same accountability system it uses to measure school and LEA progress in educating all students.

During the transition period, while states are developing their assessment programs, the yearly progress of schools and LEAs must be mea-

[1]Final Regulations §200.5(a)(2)(iii)(C) and §200.6(a)(2)(iii).

sured; however, the process may be different from the one used after final assessments have been developed or adopted.

DEFINING ADEQUATE YEARLY PROGRESS

States must develop procedures, or approve locally developed procedures, to determine whether schools and LEAs are making adequate yearly progress (AYP) in enabling children served by Title I to meet the state's student performance standards. While states may allow the use of locally developed procedures, states are responsible for ensuring that these procedures are appropriate and for defining what constitutes adequate yearly progress.

The regulations (§200.3(b)) require each state to define adequate yearly progress in a manner that

1. Results in continuous and substantial yearly improvement of each school and LEA sufficient to achieve the goal of all children served under this subpart, particularly economically disadvantaged and limited-English proficient children, meeting the state's proficient and advanced levels of performance;
2. Is sufficiently rigorous to achieve the goal within an appropriate timeframe; and
3. Links progress primarily to performance on the state's assessment system under §200.4, while permitting progress to be established in part through the use of other measures, such as dropout, retention, and attendance rates.

Procedures used for determining AYP for schools may be different from those used to determine whether LEAs are making AYP. For example, LEA progress measures may be based on the proportion of schools making adequate progress.

Student performance, as measured by the state's assessment system, is the primary source of information for determining whether schools and LEAs are making AYP. If a school does not participate in the state assessment system (e.g., a K-2 school in a state with an assessment program that begins in grade 3), other appropriate measures must be used to gauge the school's progress.

In targeted assistance schools, measures of progress may be based on the performance of all students, the performance of only students who are currently served by Title I, or on the performance of students who are being served and have been served by Title I. In schools with schoolwide programs, the performance of all students is used to deter-

mine the school's progress. LEA progress may be based on the performance of all students or on the performance of only students who have been or are served by Title I.

In measuring progress, states may choose to include the performance of students who have been served by Title I but are not currently served. This option can be useful in determining the progress of a school in which most of the students who are served by Title I are served in non-assessed grades. For example, a K-5 targeted assistance school may serve students primarily when they are in grades K-3. If the formal assessment program begins in grade 4, the performance of fourth-graders who had been served under Title I in K-3, as well as those currently served in grade 4, might be used in determining progress.

Another option for determining school progress is to include assessment data from the receiving school in the feeder school's evaluation. For example, student performance on a grade 6 assessment could be part of determining the progress of the feeder K-5 school.

USING INFORMATION ABOUT PROGRESS

AYP results are used to identify schools and LEAs in need of improvement. Schools and LEAs that have not made adequate progress for two consecutive years are identified as needing improvement (schools have an opportunity to present information that contradicts the designation). Once a school or LEA is identified as needing improvement, procedural and programmatic assistance begins to take effect.

Information about a school's progress is also used to identify distinguished schools. In most cases, a school must have exceeded the criteria for AYP for three consecutive years (and may need to meet other requirements) before being identified as distinguished.[2] States may grant distinguished schools privileges such as monetary rewards or freedom from specific regulations.

STANDARDS-BASED MEASURES OF PROGRESS

Quite often, the progress of a school or LEA is measured based on the average performance of students. Although averages can give an overall picture of achievement, they can mask more than they reveal.

[2]Schools in which "virtually all children have met the State's advanced level of student performance" and schools in which "equity in participation and achievement of students by sex has been achieved or significantly improved" are also eligible to be designated as distinguished schools (§1117(c)(2)(A)).

Changes in average test scores say nothing about how different groups of students, such as low achieving students, are progressing over time.

In the standards-based system mandated by Title I, a logical alternative is to measure progress against pre-defined performance levels: the number of students identified as performing at each level of performance. (Kentucky and Maryland are two states that use standards-based reporting to describe school performance.) Title I requires states to define adequate yearly progress (AYP) in terms of enabling students to meet the state's proficient and advanced levels of performance.

DETERMINING AYP — AN ILLUSTRATION

This section highlights some of the factors involved in determining AYP through the use of a series of examples. The examples for measuring progress outlined here use the proportion of students at each performance level as the basic unit. Changes in these proportions over time are used to gauge the progress of schools and LEAs. To illustrate the basic approach, assume that

- ✔ There are four performance levels that describe student performance: advanced, proficient, partially proficient, and novice.
- ✔ Student performance has been measured in such a way that each student can be classified as performing at one of these four levels.
- ✔ The goal is to increase the proportion of students classified as proficient or above.
- ✔ Schools are the focus (although these procedures can apply to LEAs).

Results are reviewed annually, and schools that do not meet progress requirements for two consecutive years may be put in program improvement.[3]

The five examples given below increase in complexity by building on each previous example. They are designed to illustrate some of the basic concepts in using this approach to determining AYP. The numbers used in the examples (e.g., for percentage at each performance level; for weights at each performance level) have been chosen to make the illustrations as simple as possible and do not necessarily represent numbers found in state and local programs.

[3]States may apply for waivers to modify the annual review process. Kentucky, for example, reviews data combined over a two-year period in determining adequate progress.

Five Examples of Measuring Annual Progress

EXAMPLE	CRITERIA FOR AYP
1	Increase in the proportion of students classified at proficient or above
2	Increase in the proportion of students classified at proficient or above, to meet a specified goal within a specified time frame
3	Same as Example 2, with the addition of reducing the proportion of students classified at novice
4	Same as Example 3, with the addition of increasing the proportion of students classified at advanced
5	Same as Example 4, with results weighted to reflect the importance of the changes in specific categories

EXAMPLE 1: *Increasing the proportion of students classified as proficient or above.*

In the simplest case, a school is making progress if the total percentage of students classified as either proficient or advanced increases.[4]

PERFORMANCE	BASELINE	YEAR 1
Advanced	15%	16%
Proficient	35%	39%
Partially Proficient	25%	22%
Novice	25%	23%

EXAMPLE 2: *ADD a goal and a time frame to meet this goal.*

For Title I, the ultimate goal for schools and LEAs is that all children served meet the proficient and advanced performance standards. If this goal is added to the simplest case, and a time frame for meeting this goal is determined, a school is making progress if the total percentage of students classified as either proficient or advanced increases at a rate sufficient for it to meet the goal in the given time frame. For example, the time frame for meeting the goal might be 10 years, and schools might be expected to move steadily toward meeting that goal. In this example, a school that started with the set of data illustrated below would be

[4]The examples are designed to illustrate various concepts — sample data are not intended to represent actual school performance.

expected to increase the proportion of students scoring at proficient or advanced by 50 percentage points over the 10 years to meet the goal of 100% at proficient or above. Since a steady rate is expected, the school would be making AYP if it increased the proportion at proficient or advanced by 5 percentage points annually.

PERFORMANCE	BASELINE	YEAR 1	→	YEAR 10
Advanced	15%	16%	→	30%
Proficient	35%	39%	→	70%
Partially Proficient	25%	22%	→	0
Novice	25%	23%	→	0

EXAMPLE 3: *ADD a requirement to reduce the proportion at the novice level.*

Although the ultimate goal is for all children to be performing at least at the proficient level, states may wish to define progress in a way that recognizes improvements in educating students at the lowest performance level. In this case, a school is making progress if the proportion of children at the novice level decreases and the proportion at proficient and above increases, all at a rate sufficient for it to meet the goal in the given time frame. Again assuming a 10-year time frame and steady growth, the school illustrated above could be considered making AYP if it 1) increased the proportion scoring at proficient and above by 5 percentage points AND 2) decreased the proportion scoring at novice by 2.5 percentage points.

EXAMPLE 4: *ADD a requirement to increase the proportion at the advanced level.*

States may wish to further define progress as including improvements in educating children who are performing at the proficient level, so that more children are scoring at the advanced level — that is, any increase in the proportions of students scoring at higher levels signifies progress, as long as there is a related decrease in proportions scoring at lower levels, all at a rate sufficient for the school to meet the goal in the given time frame. Assuming a 10-year time frame and steady growth, the school illustrated above could be considered making AYP if it 1) increased the proportion scoring at proficient and above by 5 percentage points, 2) decreased the proportion scoring at novice by 2.5 percentage points, AND 3) increased the proportion scoring at advanced by 3.5 percentage points.

PERFORMANCE	BASELINE	YEAR 1
Advanced	15%	18.5%
Proficient	35%	36.5%
Partially Proficient	25%	22.5%
Novice	25%	22.5%

EXAMPLE 5: *ADD expressing progress using a weighting system.*

When more than a single category of movement is included in the definition of progress, as in Examples 3 and 4, it becomes complicated to describe the progress of a school. For example, if a school's data are as presented below, has the school made adequate progress over time? The school has increased the proportion at proficient or above, reducing the proportion at partially proficient, but the proportion categorized as novice has increased, and the proportion at advanced has stayed the same.

PERFORMANCE	BASELINE	YEAR 1
Advanced	15%	15%
Proficient	35%	40%
Partially Proficient	25%	15%
Novice	25%	30%

A weighting system can attach a value to different types of change so that the interpretation is easier and consistent. A school "score" can be developed by differentially weighting the proportions in each category, multiplying the proportions by their respective weights, and summing the results. For example, movement from below proficient to proficient or above may be most highly valued, followed by movement from novice to proficient and then movement from proficient to advanced. In this case, weights might be advanced, 125; proficient, 100; and partially proficient, 30. This weighting scheme awards 30 points for movement from novice to partially proficient, 70 points for movement from partially proficient to proficient, and 25 points for movement from proficient to advanced. The goal, in terms of the score, could be set at 100 (100% x 100, the weight at the proficient level).

PERFORMANCE	WEIGHT	BASELINE PERCENT	BASELINE SCORE	YEAR 1 PERCENT	YEAR 1 SCORE
Advanced	125	15%	18.75	15%	18.75
Proficient	100	35%	35.00	40%	40.00
Partially Proficient	30	25%	7.50	15%	4.50
Novice	0	25%	0	30%	0
SUM			61.25		63.25

Using the 10-year time frame and assuming stable growth, this school would be meeting AYP requirements if its score increased by 3.875 points every year [(100-61.25)/10]. In this example, the school's score increased by only 2 points in one year; therefore, the school did not meet annual progress requirements.

In general, a weighting system such as the one described moves measures of progress away from the "pure" standards-based system. Students performing at the advanced level can compensate for students performing at below proficient levels. For example, a school could meet the goal above of 100 if its data looked like this:

PERFORMANCE	WEIGHT	PERCENT	SCORE
Advanced	125	15%	18.75
Proficient	100	80%	80.00
Partially Proficient	30	5%	1.50
Novice	0	0%	0
SUM			100.25

COMBINING INFORMATION — EXTENDING THE BASIC APPROACH

The basic approach illustrated above can be applied easily to student performance in a single content area. However, at a minimum, AYP must be defined based on performance in two content areas, reading/language arts and mathematics. In addition, states may choose to include other factors in defining progress. If only a few factors enter into the AYP decision, the basic approach may be sufficient. For example, consider a state that includes attendance rates and dropout rates in determining progress. This state's definition of AYP for a school could

use a model with multiple criteria. Adequate progress could be defined as meeting the progress requirements for both reading/language arts and mathematics and either improving the average daily attendance rate by a certain percent or decreasing the dropout rate by a certain percent. (Other combinations are possible — e.g., meeting any two out of the four criteria, meeting at least one specified criterion and any two of the other three criteria, or meeting all four criteria.)

Using an Index to Combine Information

Another way to combine information is to use an index. An index can be used with any appropriate weighting scheme (including one that rewards only progress from below proficient to proficient and above). A major advantage of using an index is that it provides an easily understood summary of a complicated combination of information (the Dow Jones index is an example). A major disadvantage is that an index is not easily interpreted beyond a global level, since it may combine a variety of types of information.

The best way to illustrate the use of an index system is via a simple example. An elementary school serving grades K through 5 has a schoolwide program. School progress is defined by the following components:

- ✔ Performance in reading/language arts and mathematics in Grade 2, based on an LEA portfolio assessment system
- ✔ Performance in reading/language arts and mathematics in Grade 5, based on the state on-demand assessments and the LEA portfolio assessment system

These components could be weighted as follows, so that Grade 2 performance "counted" as much as Grade 5 performance.

COMPONENT	WEIGHT
Grade 2 reading/language arts — portfolio	4
Grade 2 mathematics — portfolio	4
Grade 5 reading/language arts — state assessment	2
Grade 5 mathematics — state assessment	2
Grade 5 reading/language arts — portfolio	2
Grade 5 mathematics — portfolio	2

Each component would be given a score, as illustrated in Example 5 (page 57), each score would be multiplied by its weight, and the results would be summed. This sum would then be the index score for the school. The school index would be compared to the school's goal for the period to determine whether the school made adequate progress.

Index systems can be relatively simple or as complex as needed, incorporating several transformations of information. For example, goals could be set for each component, and scores for each component could be transformed into percent of goal met. Percents of goal met could be transformed into scores that are subsequently weighted and summed to produce an index score. Care must be taken in selecting data to be included and in determining how data are weighedt, with more care needed the farther the index strays from the original data sources.

The paper, *Adequate Yearly Progress in Title I of the Improving America's Schools Act of 1994: Issues and Strategies*, describes in depth a number of variations on this model for determining progress. Examples contained in the monograph illustrate key issues to be considered in developing an approach for measuring school and LEA progress.

New Challenges and a Unique Opportunity

Although Title I imposes some general requirements on how progress is measured, states are responsible for making most of the decisions involved in developing processes for measuring AYP. The time permitted for states to develop their final assessment plans — and the transitional assessments that they can use until their final assessments are ready — offers states the opportunity to make key AYP decisions in a careful and considered manner. Some of the questions states can use this period to address are described below. The issues involved are often complex and interrelated; how each state answers these questions will depend, in large part, on the nature of its standards and assessment system. Lessons learned from past practice and current research can help guide states in their decision-making. States can also learn from the experiences of other states that have addressed or are in the process of addressing these questions (e.g., Kentucky, Maryland). Close cooperation between states and local educational agencies will be necessary to develop a program that serves needs at the school, local, and state levels.

If a state allows the use of local standards or assessments, how will the state ensure that they are appropriately rigorous?

Allowing LEAs to use locally developed standards and/or assessments for Title I purposes does not relieve the state of its responsibility to have a system (in this case, a system that includes local components) of rigorous content standards, performance standards, and assessments. How will states review local standards and assessments? What criteria will be used to determine whether the content and rigor meet the state's expectations? Will models be developed for local use? How will the state ensure that standards and assessments are aligned at all levels? How will progress be measured equitably if the state has several different locally developed systems of standards and assessments? What kind of technical support will the state provide? How can LEAs work together on developing standards and assessments?

What should the time frame be to reach the goal of Title I?

States must set a reasonable time frame for schools and LEAs to meet the Title I goal: all students performing at the proficient or advanced level. Is 15 years too long? Is five years too short? In addition, states must determine the rate and nature of school and LEA improvement requirements. Should schools be expected to improve at a steady rate over time, or should improvement be faster in the beginning years? Should the rate of improvement depend upon where the school or LEA starts?

What information should be included in measures of progress?

Measuring progress of schools and LEAs is essentially an evaluation task. Decisions must be made about what information best describes progress in enabling students to meet high standards. Will performance in academic areas other than mathematics and reading/language arts be included? Will non-academic indicators (e.g., attendance) be used? How will information about specific student populations (e.g., students with limited proficiency in English, economically disadvantaged students) be used in determining whether a school or LEA is making AYP? Will the system have a separate definition of progress (e.g., requiring a faster rate of growth) for these students to insure that substantial progress is made? How will the academic progress of grade levels not covered by the assessment system be measured? Will the achievement of students who have been served but are not currently served by Title I be included?

Will the system be compensatory or conjunctive?

Progress may be measured using a compensatory system, in which high performance on one criterion may make up for low performance on another criterion (e.g., in a system that combines information from var-

ious assessments, relatively low reading performance might be compensated for by relatively high mathematics performance). Alternatively, a conjunctive system, in which schools and LEAs are required to meet each in a series of criteria, may be used (e.g., in a system that considers each assessment separately, schools are required to meet a particular standard on both the reading and mathematics assessments). What are the effects of using each kind of system? Is a conjunctive system too harsh? Does a compensatory system allow some schools to fall through the cracks or allow low performance on an important criterion to be counteracted by high performance on a less important criterion?

How will the state assessment system be used in measuring progress?

A number of characteristics of the state's assessment system will affect how AYP is measured. How are reading/language arts and mathematics assessed — at what grade levels and using what measures? Will the same assessment be used to report individual results and to provide data for measuring school- and LEA- level progress? Will some form of sampling be used in the assessment design? How will information from multiple measures be combined?

How will students with disabilities be assessed?

The assessment system must be appropriate for all children, including students who have often been excluded from the assessment system in the past. How will assessments be modified for students with disabilities? How will results of modified assessments be reported? Can a system be developed so that modifications and accommodations are not necessary?

How will students with limited English proficiency be assessed?

Limited English proficient (LEP) students often are exempted from the assessment system for an arbitrary period of time (on the assumption that during that time, students will become proficient in English). The law requires that LEP students served by Title I be included in the assessment system, to the extent practicable using the language most likely to yield accurate information about student performance on standards in subjects other than English. How will appropriate accommodations be designed for LEP students so that they can participate in the assessment? How will the determination of the type of accommodations necessary for individual students be made? How will results of modified assessments be reported? Can a system be developed so that modifications and accommodations are not necessary?

How accurate is the AYP determination?

Title I allows for inaccuracy in the decisions that are based on progress measures. If a school or LEA feels it has been misclassified as needing improvement, it may provide evidence to dispute that classification. How accurate do the classifications need to be? How can instability in the student population be accounted for? Is two years long enough for accurate results? Does a compensatory system provide more stable results than a conjunctive system? What are the effects of errors of measurement when multiple measures are combined? What system of local appeals will the state develop?

How can the system for AYP contribute to improving student learning?

Although whether a school or LEA is making adequate progress is a yes/no decision, the data that go into making that decision should be data that can tell schools much more about how their students are performing. How will the information that is used to determine progress for Title I purposes be connected to other information about students? What kinds of information will schools be given to help them understand their AYP results? What support do schools need to use their data well?

How can disaggregated results be reported appropriately?

Assessment results must be able to be disaggregated at the school and LEA levels and be reported (if statistically sound) by a variety of classifications. How should "statistically sound" be defined? How can results for subpopulations be productively used if they are not statistically sound? Are there descriptive techniques that would help make these results meaningful?

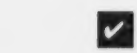
Summary

Title I requirements for Adequate Yearly Progress are intended to encourage and support high quality education for all students. Differences in Title I requirements from past requirements make designing an appropriate system particularly challenging. This report has provided an overview of AYP requirements and outlined steps states can take and issues states should consider in designing their systems. The second paper on Adequate Yearly Progress in this series, *Adequate Yearly Progress in Title I of the Improving America's Schools Act of 1994: Issues and Strategies*, contains explicit strategies and examples that states can draw on to implement Title I AYP requirements.

HOW SHOULD THE COMPREHENSIVE ASSESSMENT SYSTEM BE DESIGNED?

TOP DOWN? BOTTOM UP? BOTH? NEITHER?

by Edward Roeber



COUNCIL OF CHIEF STATE SCHOOL OFFICERS
WASHINGTON, DC

Introduction

Much discussion is taking place in Washington, D.C. about the comprehensive assessment systems which the nation needs in order to encourage the types of teaching and learning which the nation's students need in order to learn the types of skills and behaviors required in later education and work. This certainly is the golden age of assessment. Unlike any other time in the nation's history, student assessment is viewed as the pivotal piece around which the reform and improvement of the nation's schools turn. Assessment designs are being developed at the national level, in many states and at the local level (in many local districts, as well as through discussions about the future of Chapter 1 evaluation designs for local school districts). Assessment of student performance is viewed as the means to set higher, more appropriate targets for students, as well as the means to focus staff development efforts for the nation's teachers, the mechanism to encourage curriculum reform and the vehicle for improvement of instruction and instructional materials in a variety of subject matters and disciplines. Student assessment carries a heavy load these days!

An essential element in these discussions which is just beginning to emerge among those discussing student assessment is how the assessment program designs which are under discussion at the national, state and local levels will "fit" together to make up a comprehensive assessment system. By "comprehensive," the author means a system which serves the purposes of the assessments proposed at the various levels in a manner which is coordinated across the various levels so that duplicative data collection is eliminated and yet multiple sources of information are used, especially when important decisions are being made about students, schools or instructional programs.

Heretofore, the plans being proposed at the national level have not featured this coordination. Each has been proposed either independently from other assessments currently taking place or stuck on to those, with perhaps the vague hope that some of that other "stuff" would simply disappear. States, trying to make sense of what is occurring nationally and to rationalize this with what they are doing, often have put these national proposals together with various current and proposed state efforts in a crazy-quilt of programs and purposes and in a manner which calls for more assessment than is needed or can be used. Their plans may be so broad that all proposals for assessment at the national and the

state levels are "in there." The result may be too much testing of students and an angry backlash of sentiment from teachers and others at the local level against all of the assessment efforts.

At the same time that student assessment has emerged at the top of the list of things to tinker with at the national and state levels, concern about the form of assessment has also been growing. As the nation becomes increasingly concerned about the skills which the nation's students possess (or do not possess), an essential element of the acquisition of these desired skills is the form of assessment used in this country. Unlike other countries, the United States relies almost exclusively on the multiple-choice test as the means of assessing student performance. Critics indicate that such assessments tend to stress basic knowledge types of skills and tend to encourage teachers to stress memorization of content, rather than the use of content to solve meaningful problems.

A variety of content-area groups are currently re-examining what they view as important and how schools should be teaching these outcomes. A common element across these subject-matter groups is the de-emphasis of content knowledge and an emerging emphasis on application and use of the content, even to the point of consciously deciding not to teach certain content. The current philosophy is that "less is more," that it is more important to study key questions and issues in depth and to be able to have thoughtful discussions about this material which lead students to informed decisions. These redefinitions of content areas, now currently under way in the efforts to set national content standards in areas such as science, history, geography, the arts and English/language arts, are all featuring this shift in what is considered important for what students need to be able to do.

It is this growing shift in emphasis in student outcomes which is leading some at the national, state and local levels to emphasize new means of assessing student performance. Emerging techniques, such as the use of portfolios, projects, exhibitions, demonstrations, individual performance assessments, group performance assessments, hands-on assessments, and so forth are increasingly being used as the most appropriate, most-directly-tied-to-instruction techniques for assessing students. It is ironic that these techniques, plus others (such as use of anecdotal records, observation, structured interviews and others) which are widely viewed as most suited for classroom-level assessment, are emerging as potential strategies to be used in large-scale assessment programs. Policy makers and others view these instruction-related assessment strategies as the ones to use for assessment programs tied to instructional improvement.

Therefore, as the policy makers, assessment specialists and curriculum specialists look to design the rational and comprehensive assessment system needed by this country, important elements of this discussion include what skills are most critical, how will these be learned and what are the assessments that will inform and lead students to this learning? What is the assessment design most appropriate for the nation, the states and the local districts, and should this design focus on assessment procedures traditionally used in large-scale programs, use assessment strategies which have been used predominantly at the classroom and school level, or both? This is an important issue for the nation's schools.

Purposes for Assessment

Before discussion about the design of a comprehensive assessment system can take place, it is important to consider what are the purposes of student assessment at the student, school, district, state and national levels. These purposes can be summarized as follows:

MONITORING

- ☑ *Student Level:* Provide periodic measurements of student progress in order to determine the educational "growth of a student from year to year."
- ☑ *System Performance:* Provide a periodic measure of the performance of groups of students to track performance over time.

INFORMATION/ACCOUNTABILITY

- ☑ *Parents and Students:* Inform parents/student about student performance so as to encourage student or teachers to improve performance.
- ☑ *Public:* Provide the public with information about the performance of groups of students so as to encourage schools to improve the system.

IMPROVING STUDENT PERFORMANCE

- ☑ *Student Level:* Provide data to teachers and students which encourages instruction geared to the needs of individual students.
- ☑ *System Level:* Provide information to educators on groups of students which can be used to review current instructional strategies and materials and make improvements where needed.

ALLOCATION OF RESOURCES

- ☑ *Human:* Use information to determine where instructional staff are needed.
- ☑ *Financial:* Determine where financial resources are most needed.

SELECTION/PLACEMENT OF STUDENTS

- ☑ *Selection:* Help determine the eligibility of students for various educational programs or services.
- ☑ *Placement:* Determine the program or service most appropriate for the instructional level of the student.

CERTIFICATION

- ☑ *Student Level:* Provide a means of determining the competence level of individual students.
- ☑ *Program Level :* Provide data to certify the adequacy of an educational program.
- ☑ *System:* Provide data to certify the acceptability of an educational system (Accreditation).

PROGRAM EVALUATION

- ☑ Provide the information needed to determine the effectiveness of an educational program or intervention.

Although there are a variety of ways in which to describe the purposes of assessment, this list provides an overall view of the variety of purposes for student assessment, as well as the manner in which programs may have multiple purposes.

Another way to view this is depicted in figure 1 below.

FIGURE 1 COMPREHENSIVE ASSESSMENT SYSTEM — PURPOSES FOR ASSESSMENT

ASSESSMENT LEVEL	REMEDIATION	SELECTION/ PLACEMENT	IMPROVE INSTRUCTION	CERTIFICATION OF STUDENT	MONITOR/INFORM SCHOOL	HOLD ACCOUNTABLE	ALLOCATE RESOURCES
Internal						X	
National						XX	X
State			X		X	XXX	XX
District		X	X	X	XX	XX	XXX
School	X	X	XXX	X	XXX	X	X
Classroom	XX	XX	XXX	XX			
Student	XXX	XXX	XXX	XXX			

Note: Each X indicates the anticipated match between purpose and level of assessment, assuming that appropriate assessment instruments are used.

Formats for Assessment

The design of the assessment program to be used is an element of assessment which needs to be carefully thought out. Until recently, most assessment programs used primarily multiple-choice tests, with essays used occasionally where needed, such as in writing examinations. Although programs such as the National Assessment of Educational Progress pioneered the use of a broader array of assessment formats, much of that innovation was dropped for budget reasons fifteen years ago and hence, it does not serve to model other possible designs.

Because the nature of the skills which educators consider to be important is changing, the manner in which the skills are assessed needs change as well. As educators and others change from an emphasis on the acquisition of content knowledge to an emphasis on the use of the knowledge to apply to unique problems, the means of assessing students has been changing. In order to see if students are capable of defining an issue, determining what information is needed to respond to the issue and to provide a rationale for their response to the issue, students will need to be assessed in a way which allows them to set up the problem, select the information needed and provide their unique and extended response to it. Multiple-choice items simply don't allow such assessments to occur. Alternatives are needed.

These alternatives are increasingly being used in large-scale assessment programs. Such alternatives as the use of essays or extended response formats in assessing areas such as mathematics or reading, the use of interview questions, hands-on performance exercises, group performance exercises which are scored both for end result and group process are some of the "on-demand" types of exercises now finding their way into large-scale assessment programs. Some programs are now using a few of these types of exercises, while a couple of states' programs are now based solely on these alternative programs.

In addition to these on-demand types of exercises, other assessment formats which break out of the time constraints and allow students to develop a response over time are currently being used in a few large-scale programs. Because some of the types of thoughtful discourse which we wish students to engage in simply can not be assessed in a few minutes or an hour, since they may require a student to gather information, synthesize the data into that most pertinent, organize it in a convincing manner and present it succinctly, there have been attempts to break the bounds of on-demand types of assessments. Such assessment formats include the use of portfolios, demonstrations, exhibitions, and

projects. Although these are different ways of assessing students, they are similar in that they allow students to prepare the assessment task over a period of weeks, months, even years. Although they can serve to document the process of responding to an assessment task, their primary advantage is that they allow students to engage deeply in responding to an assessment task. They can also allow students to select the task for assessment themselves (that is, students can select the project topic, so it is one which they are interested in and have some background with which to respond).

As the large-scale assessment program is being designed, both the purpose(s) for assessment, as discussed in the previous section, and the types of assessment formats to be used, must be considered. Some types of assessment instruments are listed below.

Types of Assessment Instruments

ON-DEMAND ASSESSMENTS

- ☑ Multiple-choice exercises
- ☑ Short-answer, open-ended
- ☑ Extended-response, open-ended
- ☑ Individually administered interview
- ☑ Individually administered performance events
- ☑ Group-administered performance events

EXTENDED-TIME ASSESSMENTS

- ☑ Individual performance tasks
- ☑ Group performance tasks
- ☑ Projects
- ☑ Portfolios of student-selected tasks
- ☑ Observations
- ☑ Anecdotal records
- ☑ Others

Information Needed at Different Levels

It is important to consider the information about student achievement needed at various levels in the educational system. Certainly, the information needs of the parent are different from that of the teacher; the parent wants to know what the student can do and not do, while the teacher is more concerned with what are the student's strengths (what aspects of the learning task has the student accomplished) and what additional work is needed. These information needs are different from that of the building principal, who needs to know if achievement in the school is comparable to that elsewhere and more broadly, whether students are learning what they need to learn.

At the district level, the concern may be more whether the achievement needs are greater in mathematics than reading, so that additional resources (extra staff and additional instructional materials) should be allocated for mathematics instruction. At the state level, the concern is often whether there is equity in school programs and whether differences in student performance are due to the lack of resources in some schools; issues of funding, incentive programs for improving in key areas such as mathematics, staff training and so forth are discussed on a global level. Underlying these discussions is a concern about whether the students in the state are economically competitive in a manner comparable with other states and other nations.

This competitive concern also permeates the discussions at the national level. Although policy makers at this level are concerned about differences between states, as well as types of school systems, the underlying worry is about how much American students are learning in comparison to their peers in other countries, particularly those which are our trading partners.

These information needs, which can be seen to be very different from the teacher level to the national level, form the basis for the discussions about assessment design. In the top-down models, assessments which meet the needs of policy makers at the state or national levels are developed and implemented, but because we presume that the data may be useful for other purposes, we take this data and try to convince others, such as building principals, teachers and parents, that the information will really meet their needs as well. An emerging alternative to this is to build the assessment system which teachers, parents and students need and presume that the users at the district, state and national levels can have their questions answered just by aggregating the types of assessments used at the classroom levels.

Is it feasible to design one assessment program which can meet all needs? Serious questions have been raised whether any one assessment program design can serve all purposes. Yet, many assessment specialists are seriously engaged in seeking the one design which will meet all of the needs. An alternative approach is needed.

School Improvement Model and Assessment Information Needs

An alternative approach to assessment program design is to break out of the assumption that one assessment program will meet everyone's needs. Based on an alternative assumption that multiple approaches will allow different users' needs to be met, a more comprehensive assessment design can be developed. This assessment design can use different assessment formats to meet the needs of different users.

A second key assumption is that because of the nature of each assessment format, measurement error is introduced with each type of assessment format used. Although the shortcomings of the multiple-choice item are well documented, the alternative assessment procedures introduce their own biases as well. If this is true (and the limited evidence on comparisons of traditional and alternative assessments strongly suggests that it is), then a more fruitful approach will be to consider using an array of assessment formats in the assessment design. This will help answer the nagging question of just how much of a student's achievement or lack thereof is due to the assessment format and how much is "true" achievement.

An important place where the multiple purposes of assessment and the multiple formats for assessment come to bear is at the building-level school improvement planning process. Although there are several variations on a theme, it is typical that the staff of the building (or their representative) work together to develop the school's mission, goals and long- and short-term objectives. Each of the objectives is measured in order to determine the extent of accomplishment of the school's goals (and, hence, the school's mission).

Typically, the objectives of the school improvement team are measured in one way; objectives dealing with student achievement are measured with available student test data, often times the information from the testing program implemented to meet the needs at the district or state levels. Information collected formally and informally at the classroom levels is not typically included in such school improvement plans, even

though such information could provide an alternative approach to measuring the same types of objectives and goals.

Using the assumption of multiple formats for measuring student performance, as well as the different assessment information needed at different levels resulting in different approaches to assessment, an alternative model for assessment could be designed.

Top Down or Bottom Up? Both!

One important use of information at the building level is within the school improvement plan of the building staff. The approach to the design of assessment systems becomes most obvious at this level. Hence, it is important to discuss where data to be used at these levels comes from.

The school improvement planning process which is becoming increasingly popular across the country is designed for the building staff to determine the school's mission, long-term goals and short-term objectives, to measure these objectives and to determine how the school can become more effective for all students. Within this system, information on the achievement of students is vital. Does this information come just from the assessments external to the school, those imposed by the district, the state or the nation? Is it feasible and desirable for the information which individual teachers and groups of teachers gather in their classrooms to be used? If so, is this in place of the external measures?

The tendency of local educators in designing school improvement systems is to attempt to locate one measure for each objective they list in their school improvement plan. If the improvement of reading achievement is an objective for the school, then they debate whether to use the state's reading test or the teacher's reading portfolios as the sole measure of the accomplishment of this objective. A far more fruitful approach for the building team to consider is to select two, three, four or more different measures for each objective they list in their school improvement plan. Rather than debating whether the state test or classroom portfolios are the best measure, use both of these and perhaps others as well.

In addition, when these multiple measures are used, the team should resist the temptation to select the one measure which they feel is most accurate to report accomplishment to the parents and the public. Instead, the building team should report all of the results. Will they contradict one another in some instances? Yes, they probably will. Is

this important to present? Yes it is, because it will help to inform parents and the public that the information yielded by tests may be different than that yielded by classroom observations, interviews, performance measures, portfolios and so forth. Does different results mean that one measure is "valid" and another "invalid"? It may, but more likely, it means that the measures were assessing students differently, either because they assess different outcomes within an area like reading or because they differ in format and content, students perform differently on them. Reporting all of the measures and displaying all of their differences is one way to show that the format and content of assessment procedures does have an impact on assessment information!

The Ideal Comprehensive Assessment System

Beyond the building-level needs for information, it is feasible to gather information at the classroom level and aggregate it upwards, but is it the most cost-effective instrumentation to answer the questions posed by state-level policy makers concerned about the performance of school systems within the state? Does the instrument which answers these policy makers questions (which may be a multiple-choice survey instrument) serve to effectively inform classroom teachers about the learning growth of an individual student? The portfolio of work which would be much more effective for this teacher may not be cost effective in answering the state-level questions. Given the different information needs of persons at different levels of the educational system, plus the variety of reasons why assessment takes place and the various tools available for assessment, is it any wonder that one instrument imposed at all levels will not meet all the needs and purposes for assessment?

It is virtually impossible to meet these different needs and purposes with a single instrument and to do so in an efficient and effective manner. What is needed then is a top-down and bottom-up approach to assessment. In one ideal assessment system, some data will be collected at the national and state levels, some information at the state and school district levels, some at the school district and building levels, some at the school and classroom levels, some at the classroom and student levels and some only at the student level. While this type of system is a bit extreme, since it may occur that a common question could be addressed with a common instrument given to samples of students at the international, national, state and perhaps even local district levels (as has been proposed for the IEA Third International Mathematics and Science Study, particularly if it is linked to the National Assessment of Educational Progress' mathematics and science assessments), the model

does serve to point out that multiple instruments may be preferred in an ideal assessment system. In addition, by presenting assessments which do not span the full range from the student to the international level, the model is also combining top-down and bottom-up approaches to assessment. Both are useful considerations to keep in mind for designing the assessment system which best meets the variety of needs for achievement information.

The ideal assessment system is one, then, which has identified the various audiences for assessment information, the types of information which each audience needs (keeping in mind that each audience may have the need for more than one type of information), the type of assessment instrument which best meets the assessment need, the impact of the use of the instrument on the educational system, and the levels for use of the assessment instrument. Undoubtedly, both conventional types of measures and those emerging in popularity, such as open-end exams, portfolios, projects, performance assessments and so forth, will be used in the system. Both conventional and the emerging measures have advantages and should be used where the measurement need is such that the type of measure can be used to the greatest advantage.

While cost-effectiveness has been mentioned previously, this should not be just a formula for estimating how to gather data most economically. While the cost of gathering data is important, the "cost" of using or not using particular assessment instruments also needs to be calculated. These types of costs are much more difficult to estimate, but through surveys of teachers and classroom observation and interviews, it is possible to estimate the impact (or lack thereof) of the assessment procedures which are used.

Components of a Comprehensive Assessment System

DOCUMENTATION SYSTEM

- ✔ Portfolios to record student work
- ✔ Inserts for the portfolios about parent conferences, teacher observations, student comments and so forth
- ✔ Places to record teacher comments on student work
- ✔ Places to record student self-reflections
- ✔ Assessment tools and resources
- ✔ Proficiency scales derived through observation

- ☑ Behavioral rating scales for student attitudes and behavior
- ☑ Structured and unstructured observations in both staged and naturalistic settings
- ☑ Interview protocols
- ☑ On-demand, disciplinary and interdisciplinary assessments using selected response, short-answer and extended response formats
- ☑ On-demand, disciplinary and interdisciplinary performance events (group and individual)
- ☑ Projects and exhibits
- ☑ Self-reflection/metacognitive measures
- ☑ Records of extra-classroom behavior
- ☑ Assessment protocols for judging opportunity to learn
- ☑ Records/checklists for judging student attitudes
- ☑ Others

SCORING SYSTEMS

Scoring systems are needed to assist educators and others to make appropriate judgments about students' achievement and instruction. For each type of assessment resource listed above, scoring rubrics or guides are needed.

At the *classroom level*, these will assist teachers, parents, students and others to answer questions such as these:

- ☑ To what extent is the student making satisfactory progress? How should instruction be informed?
- ☑ What are the strengths of the student, and what are the instructional implications of these?

At the *school, district or state level*, these rubrics or guides will assist educators and others to answer questions such as:

- ☑ Are students attaining the desired outcomes in the X grades?
- ☑ To what extent are students making satisfactory progress in the X grades?

REPORTING SYSTEMS

As the student work is accumulated and efforts are made to "score" it in order to answer the types of questions listed above, some effort will be needed to translate the answers for individual students gathered from the variety of assessment resources listed above into reports which integrate the information gathered from the various sources into reports which are technically sound and understandable to students, parents, and other educators. In addition, reports for groups of students are needed which summarize the information for classrooms, schools and districts.

Summary

A variety of reasons and purposes for assessment exist, as well as a variety of audiences for assessment information at the school, district, state, and national levels. Designing the ideal assessment program, the one which meets the different needs of persons at these various levels, consists of matching the purposes with appropriate types of assessment instrumentation. The full variety of measures should be considered when designing the system because some types are better suited for some purposes than others. An illustration of the school improvement team was used to convey that there is a strong advantage in using multiple measures at the level of the school building to collect information on the success of the school in meeting the objectives that the staff has developed. Generalizing from this, the ideal assessment system would attempt as much as possible to capture information from the bottom up while adding measures from the top down so that users at all levels may have more than one approach to assessment to use when reporting the status of the system at that level. This approach has the advantage of demonstrating that different approaches to assessment may yield different results and that these differences help to inform the debate on what skills students should be able to perform and whether or not all students are currently able to demonstrate performance on these.

GUIDELINES FOR THE DEVELOPMENT AND ADMINISTRATION OF PERFORMANCE ASSESSMENTS IN LARGE-SCALE ASSESSMENT PROGRAMS

by Edward Roeber

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
WASHINGTON, DC

Preface

These Guidelines were first developed in 1989 for use in training workshops for state and local educators to demonstrate the processes by which performance assessments could be created, validated, and used in statewide assessment programs. At the time, virtually all large-scale assessment programs consisted of paper-and-pencil, group-administered assessments. The goal of the paper was to demonstrate that it was feasible to add to these assessments those that required special materials or administrators, so that the full range of student expectations could be assessed and reported on, at least at the state level.

However, in the time intervening since the paper was originally written, and the work that it describes was carried out, performance assessment as a methodology has been "discovered." The critics of current assessment strategies, those developing higher student standards, and those seeking simple ways of improving schools have advocated the use of performance assessments as the means of single-handedly improving school or student performance (sometimes without accompanying school improvement strategies). These individuals decry the use of multiple-choice assessments and indicate only those programs that are "pure" performance assessment based are worthy.

As ardently and fervently as the proponents expound upon the virtues and magic-bullet qualities of performance assessments, however, the detractors are equally vociferous in proclaiming the impracticality and poor technical qualities of such assessments. They point out the very real extra testing time, added cost, lack of "objectivity" of the scores, thereby attempting to convince others that multiple-choice assessment programs (which they happen to manage) are sufficient for assessing all important outcomes at the state (and perhaps local) levels. They refuse to consider whether multiple-choice and even open-ended assessments would benefit from the addition of other assessment options and ideas.

This paper is based on prior work carried out by the author in both the National Assessment of Educational Progress (NAEP) and in statewide assessment programs. The work on which this paper is based is real. Indeed, the NAEP assessments, conducted in samples of schools and homes over two decades ago at a reasonable per pupil cost, are perhaps the best demonstration that a mixed assessment program, combining multiple-choice, open-ended, and performance assessments, is both

practical and feasible. The state-level work described herein further demonstrates that it is possible to conduct valuable statewide sampling performance assessments for little cost.

This paper is offered, therefore, in the hope that the rhetorical wars can end, that the advantages and disadvantages of performance assessment can be viewed more objectively, and that such assessment methods can be used more frequently in large-scale assessment programs. The intent of this paper is to demonstrate that performance assessment is one of the important tools available for use in measuring student performance at the state or local level.

The author is grateful for the peer reviews that he has received on the original and revised versions of this paper. The reviewers challenged the incomplete thoughts and smug assumptions written into previous versions. The paper has benefitted from their critical review of it. The opinions expressed in this paper are those of the author, however, and are not necessarily those of the organization for which he works nor agencies that use or reprint this article.

Introduction

These Guidelines are provided to offer guidance to policymakers and state assessment directors concerning some of the issues in the development, administration and use of performance assessments in large-scale assessment programs. Just as it is not possible to devise the best way to develop or administer an assessment in general, it is not possible to do so for this more specialized type of assessment. However, many of the issues that builders and users of such instruments should consider are known. It is important to consider these issues as decisions are made relative to the scope and types of assessments to be used. Alternatives are available depending on the level of resources that are available, the areas in which assessment will occur, the staffing present and so forth. The purpose of these Guidelines, then, is to help the user of performance assessments consider in advance some of the issues to be faced and help plan the manner in which such assessments will occur.

Before embarking upon a description of the procedures for developing and using performance assessments, it is important to consider what the term "performance assessment" does and does not mean as used in this paper. It is not uncommon these days to hear the term performance assessment used to describe almost any non-multiple-choice assessment: short-answer questions, essays, hands-on assessments with manipulatives, group activities in which students are observed and rated

individually or as a group, and even responses for which students are given extended periods of time to respond. While the most frequently used non-multiple-choice assessment is the essay assessment in the area of writing (and such assessments do constitute a writing *performance* assessment), the term performance assessment as used in this paper is reserved primarily for those assessments that might be developed that go beyond paper-and-pencil, group-administered assessments, whether multiple-choice or open-ended. While the multiple-choice and the open-ended exercise format are valuable tools for the large-scale assessment programs, the purpose of this paper is to describe assessment methods that go beyond these means of assessing student achievement.

I. Pre-Assessment Development Activities

Before assessment development can or should occur, there are several important planning activities that set the stage for the assessment. These steps take place at the outset so that the assessment is developed in a manner that fits the content area to be assessed and is within the resources available. The planning activities described here presume that the agency conducting the assessment program has already defined certain parameters, and therefore, this paper presumes that certain parameters have already been defined. It is presumed that the agency sponsoring the assessment will have already determined:

- ✔ the assessment framework. It is assumed that the assessment framework calls for assessment strategies other than multiple-choice and open-ended. If the framework does not call for the use of performance assessments, there may be little reason to develop them.

- ✔ the grade levels at which the assessments will take place. Performance assessments at the elementary level may pose tasks or use response formats that are beyond students. At the older grade levels, student apathy may pose challenges in using such assessments.

- ✔ the purposes for which the overall assessment program, as well as the performance assessments to be developed, will serve. Are these high-stake or low-stake uses, both for students and for the professionals that work in the system? High stakes uses of the assessment will require more rigor in the development, validation, the administration, and the scoring of the assessment. In some high stakes situations, such as high school graduation tests, it may not be feasible to use performance assessments. In low stakes assessments, it may be desirable to assess only samples of students

statewide, or in a sample of schools. The stakes accompanying the assessment will help determine the design of the assessment.

- ✔ whether the results will be reported and if so, how reports of results will be provided. Will the results be combined across the performance assessments? Will the performance assessment results be combined with other types of assessment results (multiple-choice or open-ended assessments) and reported in an over-all fashion? If so, there are significant technical issues that would need to be considered.

- ✔ to whom and how the results will be reported. Will the student, teacher, school building, or school district, receive results, or will just statewide results be reported? Will group results be reported to the public? Will individual results be reported to parents? What statements about the results are to be made based on these assessments? The answers to this set of questions and the previous set will help determine the number of performance assessments to be used, how they might be scaled, and the level of detail needed for reporting.

- ✔ whether the results will be compared from year to year. The reporting of results longitudinally requires the equating of different forms of the assessment so that comparisons can be made across years, an added technical requirement.

Once these issues have been determined, the program is ready to embark on the development of the performance measures. While this Guide presents the development ideas rather informally, the more "high stakes" the program, the more it will undoubtedly be necessary to make certain that the technical, policy, and other issues have been formally addressed.

The following sections present information about how performance assessments can be developed, administered, scored, and reported. While the paper suggests informal, less costly means to develop and use these types of measures, keep in mind that some of the uses outlined above might require the use of external contractors or technical advisors.

A. DEVELOPMENT OF THE ASSESSMENT FRAMEWORK

The framework for the assessment, whether constructed specifically for the assessment or put together for more general reasons (i.e., as a document to suggest to schools the standards that students should achieve at various points of time), serves as the guide to the entire assessment. It is from the framework that the assessment is developed. Much has

been written about the process of writing performance objectives and it is not the intent of this document to duplicate that work. However, there are several ways in which such documents can be written and there are a couple of important points about such lists of objectives.

The traditional manner in which sets of objectives are written is to gather a group of content area experts and classroom teachers and ask them to indicate what students should know and be able to do by the end of particular points of instruction (e.g., by the end of third, sixth, and ninth grades). These descriptions can be few or many in number and may be briefly or extensively written up. However, the intent is to provide at least the most important outcomes and to describe these in more or less behavioral or observable terms. An alternate procedure that has been tried in some places is to allow the framework of expectations for students to evolve out of research and practical experience of educators on what outcomes students are capable of at particular times in their school career. In essence, classroom teachers and curriculum specialists explore the outcomes that students appear to be capable of and use this as the basis for the descriptions of the outcomes that students should be able to accomplish by that point. Either approach will work, with the former method probably having the advantage of both speed and lower cost, while the latter has the advantage of being more behaviorally anchored.

Regardless of the manner in which the assessment framework is developed, one important thing that must be wrestled with while it is being developed is the extent and types of performance assessments that will be used. It is possible to devise assessments that vary along the lines of the type of stimulus material to be used (audio, video or other specialized types), the types of responses (written, oral or other performance), group or individual performance, and so forth. This needs to be determined at the outset so that as the outcomes are being described, they are captured using action verbs that will be consistent with the manner in which assessment will likely occur. While it is perfectly acceptable to use action verbs that might suggest more performance assessment than is feasible (e.g., "student will describe the relationship between the post-WW I environment in Europe and the causes of WW II"), it is not desirable to describe the outcomes that would suggest a more conventional assessment (using verbs such as "select the correct answer") when a performance assessment is contemplated. Ideally, the assessment framework can be devised without consideration of the modes of assessment to be used.

B. DEVELOPMENT OF THE ASSESSMENT PLAN

Once the assessment framework has been developed, it should be rather straightforward to develop the actual assessment plan. The assessment plan differs from the assessment blueprint that will be described in the next section in both purpose and the level of detail provided. The assessment plan will provide an overview of the types of assessments to be developed and used, as well as the manner in which assessments will be carried out. The plan serves the useful purpose of describing the types of assessments that are envisioned and how such assessments will be administered, scored and reported. It is important to have this plan at the outset so that the actual assessment materials to be developed and used are known at the outset. In essence, the assessment plan serves as the introduction to the assessment blueprint.

The assessment plan should break down the assessment framework into the various types of assessments to be conducted and the types and numbers of assessment packages to be used. The plan, for example, will suggest the number of packages of the various types that are feasible in the area. In a recent paper (Roeber and Fisher, 1990) written to suggest a plan for a national assessment of the arts and foreign languages, several types of assessments were suggested. These include the following:

1. **Multiple-Choice, Group-Administered, Spiraled exercises:** these tests would not require any special apparatus (such as music tapes, works of art or videotapes) and, if individual student results are not to be reported, several such packages could be administered to different students within the same assessment group. For some packages, special printing might be needed in order to faithfully reproduce art prints within assessment booklets. If individual student results are to be reported, a common assessment will undoubtedly have to be used;

2. **Multiple-Choice, Group-Administered, Non-Spiraled exercises:** some tests that require special apparatus (music, works of art, videos and so forth) could still be administered in group settings. However, all students in a assessment group would have to be given the same assessment, since the accompanying materials apply to only one set of exercises;

3. **Open-Ended, Group-Administered, Spiraled exercises:** similar to multiple choice exercises, except that an open-end response is used in order to gather student opinions and ideas. Again, spiraled booklets may or may not be appropriate for a program, depending on the design decisions made earlier;

4 **Open-Ended, Group-Administered, Non-Spiraled exercises**: similar to multiple-choice, group-administered, non-spiraled exercises except that an open-end response is used in order to gather student opinions and ideas. These may be used where either special accompanying materials such as a videotape are used or where individual student results are to be reported;

5 **Individually Administered Performance exercises**: these exercises are designed to assess an individual's ability to produce art or demonstrate artistic skills;

6 **Group-Administered Performance exercises**: small *ad hoc* groups of students are assessed in order to determine the ability of students to work together in groups, such as improvisation in music, dance or drama;

7 **Existing-Group Performance exercises**: these exercises assess existing musical, dance and drama groups in schools that have such groups. Some schools, particularly at the elementary and middle school levels may not have such groups;

8 **Existing Student Work**: these assessments would determine whether students have, prior to the assessment time, produced works of art which can be evaluated.

As can be seen, there are a wide variety of types of assessments that can be devised to assess subject areas such as the arts or foreign languages.

The assessment plan should contain a description of the numbers of each type of assessment, as well as the number of each assessment package to be used and the anticipated assessment time. These decisions should be based on the purposes for the assessment, the manner in which results are to be reported, and the uses to be made for the assessment. Since it would be feasible to package some of the types of assessments together (e.g., the multiple-choice, spiraled exercises together with the open-end, spiraled exercises), it is important to not only estimate the number of exercises but the number of assessment packages and the assessment time per package (and total) as well. Doing this planning at the outset of the project will make obvious to all involved just how extensive the non-traditional component of the assessment will be and should go a long way to prevent the development of elaborate assessment methods that can not be used because of the lack of assessment administration or scoring resources.

C. DETERMINATION OF ASSESSMENT RESOURCES

The assessment plan should force the agency sponsoring the assessment development to carefully consider the resources needed and available in order to make decisions about the assessment at the outset. This is a critical step, one that is often overlooked until after the assessment development is completed and the agency is considering just how feasible it is to administer what has been developed. Since the postponement of such decisions about resources is likely to lead to the development of assessments that can not be used, it is far better to consider the resources that will be needed and that are available at the outset of the project. This will allow a couple of things to occur. First, the sponsoring agency can seek additional resources in order to develop and/or administer the assessment desired. Second, the sponsoring agency can change the manner in which the assessment is developed or administered (i.e., can develop or administer the assessment using volunteers rather than a paid staff or a contractor). Third, it can also change the scope of the assessment to fit within the available resources.

Several types of resources should be considered at the beginning of the project. Who will construct the assessments, who will edit these assessments, and who will develop the assessment administration and scoring procedures? Will a contractor be used, or will the assessment development be carried out by volunteers? Who will do the editing work and the preparation of materials for tryouts? What resources are available to support the development of the assessments?

Additional questions can be asked about how the assessments will be administered. Will classroom teachers administer the more conventional assessments? Will the performance assessments be administered by in-school personnel, or will outsiders need to be recruited to go to the schools to administer the assessments? Will these outsiders be individually recruited (and will they be volunteer or paid staff), or is a contractor going to be used? If volunteers are to be used, how will quality control be maintained, where it might not be feasible or permissible to dismiss a volunteer who can not competently carry out assessment administration or scoring? How will the assessment administrators be trained? What expenses will be paid? Will student responses be scored as students perform, or will they be recorded in some manner and scored later? Who will conduct the scoring, and who will participate in it? Will these be volunteers or paid scorers? Will the sponsoring agency conduct the scoring sessions or will a contractor be used for these activities?

Once, then, that the assessment plan has been written and the assessment resources have been determined, it is possible to make any adjust-

ments needed in the assessment plan. At this point, then, it is safe to assume that the plan that has been devised is a feasible one, that the assessments to be developed are known, as is the manner in which the assessments will be administered and scored. Hence, it is now possible to safely move into the next phase of the assessment development, writing the assessment blueprint. However, it should be kept in mind that the assessment plan may change during the subsequent steps in the development process, either because of unforeseen aspects to the assessment of particular outcomes or because of changes in the level of available resources. If this occurs, the agency responsible for the assessment should return to the assessment plan and revisit the answers to the above questions in order to continue to assure that the needed assessment will be developed.

D. DEVELOPMENT OF THE ASSESSMENT BLUEPRINT

Once the assessment plan has been written and accepted, it is time to develop the assessment blueprint. The blueprint will describe the characteristics of an adequate assessment for each content area of the assessment framework. Rather than focus on the assessment administration mode or the manner in which students respond, the blueprint is a description of the characteristics of the assessment for each area of the framework.

The blueprint, which should begin with a description of the answers to the questions posed on pages 81-82, would have to describe the characteristics of an adequate assessment for each student outcome. Using the example of assessment in the arts, for example, the assessment framework might indicate that students are to devise an instrumental accompaniment to a song which they hear sung to them. The blueprint for a music assessment would need to include the range of songs that a student should be familiar with at particular grade levels, the manner in which the song(s) will be presented to them (live or on a tape recorder), the manner in which students will be asked to respond, the materials to be used (e.g., stick and wood block), the manner in which student responses will be recorded and scored (scored live or tape recorded for later scoring), the criteria by which students' responses will be scored, as well as how the scores will be reported.

The blueprint will contain this information for all areas of the assessment framework. Once completed, this should guide the development of the assessments that are needed and for which resources are available. The final step in the pre-assessment development activities is for the assessment plan and the assessment blueprint to be approved by the

sponsoring agency and any additional advisory groups or individuals. While this may appear obvious, this should not be considered a trivial step, because such approval should be viewed as a commitment of resources and the sponsoring agency's will to see that the needed assessments are developed and actually used.

II. Assessment Development Steps

Assuming that the assessment plan and the assessment blueprint have been successfully developed and have been approved by the sponsoring agency and others, it is now time to move into the actual development of the needed assessment materials. It is at this stage that the initial ideas about what the assessment should contain and how the assessment will be carried out are first put to the assessment. As indicated above, it may be necessary to make changes in either the assessment plan or blueprint at this point. Such changes should be made as the assessment development process unfolds.

A. DEVELOPMENT OF ASSESSMENT PROMPTS

Given a well-defined assessment blueprint, it should be relatively easy to develop the actual assessment prompts. When developing multiple-choice exercises, this consists of writing the actual stem and responses. With open-end exercises, this consists of writing the prompt and determining at least different categories of correct and incorrect responses. With performance exercises, it is even more complex. Here the developer will need to consider how the prompt will be presented to the student, what additional stimulus materials will be needed (and, where such materials can be located or developed), how students will respond and how such responses will be recorded and scored, what criteria will be used to judge student responses, the number of scale points for scoring student responses (a four- or six-point scoring scale is typical) and samples of each level of response. Obviously, the task of devising adequate performance exercises is considerably more complex than writing adequate multiple-choice exercises.

One useful technique for developing such assessments is to do so using a combination of individual and group effort. Since there are so many aspects to the development of just one assessment prompt, it may be difficult for one individual to devise the entire exercise. However, it is also difficult for committees of exercise writers to work creatively. One strategy, therefore, is to have individuals do the initial development work on assessment prompts alone and then present their work to an

exercise writing committee. The committee can then question the individual, make editorial suggestions in the prompt, suggest additional or alternative stimulus materials, provide additional guidance on the recording and scoring of students' responses, as well as possible types of students' responses in order to refine the scoring guides. The original exercise writer or another individual can continue to refine the assessment prompt. This process can be repeated if needed or desired.

Another useful technique is to, wherever possible, try out the performance exercises with a few students during the initial developmental phase. While this may be impossible with exercises that require elaborate stimulus materials (such as specially recorded music or other stimulus materials), it is feasible with many exercises. In addition, it is most valuable if the actual exercise writer is the one who administers the exercise in this informal tryout, so that the exercise developer can see first-hand how students respond. Gathering a few student responses can assist the exercise developer in writing appropriate assessment administration directions, wording the exercise in an understandable manner and devising suggested scoring criteria and guides.

B. EDITING OF ASSESSMENT EXERCISES

Once the exercises have been written, the next step is to edit them. The editorial work for performance exercises is more complex, since there are assessment administration procedures to be followed, material to be read to the student, directions for recording student responses, scoring criteria and so forth. This is an important step in assuring that the exercise and assessment administration process are understandable, in assuring that the exercise warrants the special time and attention that performance measures require and that there is a consistency among the performance exercises.

For performance exercises, the editor will assure not only that each exercise is clearly worded, but that the "package" of exercises fit together and flow from one exercise to another. The editor can assure that the apparatus needed for one exercise will fit with the materials needed for the next one. The editor can also check to make certain that the stimulus materials fit with the exercises. Finally, the editor should assure that each performance exercise has a preliminary scoring guide that provides examples of the different levels of correct and incorrect responses, a rationale for what constitutes an appropriate response, and suggestions for reporting the responses to the exercise. Ideally, this will have been developed by the exercise writer during the exercise writing stage and may reflect informal tryouts done by the exercise writer. In almost no

case should exercises proceed to tryouts without such a scoring guide, since trying out such exercises may be a waste of valuable resources, particularly when the exercise writer and editor can think of no responses which may be judged as correct or not correct!

C. DEVELOPING ASSESSMENT ADMINISTRATION PROCEDURES

One of the more difficult aspects of performance exercises is writing the assessment administration directions. Not all performance exercises require such special directions, of course. Open-end exercises that call for students to write an essay on a topic are an example of exercises that may not require special assessment administration procedures. However, many performance exercises are administered to students individually or in small groups. These exercises require the assessment administrator to set up a standard situation for each student to respond to, as well as to read a standard set of directions to each student. In addition, the directions may also provide a standard set of probes in case the student's response is vague or incomplete. The emphasis is on the word "standard" because it is necessary to provide such standardized assessment administration procedures if the results from different students, classrooms, schools or districts are going to be combined for reporting. While again this depends on the answers to the questions posed on pages 81-82, it is something that the assessment designer should consider. To the extent that results across students are to be combined and these results publicly reported, the procedures for collecting the information will need to be standardized across the data collection sites.

In writing the assessment administration directions, there are several things to keep in mind. First, overall directions to the assessment administrator, that cover such things as drawing the sample of students within the school, locating a suitable room for testing, and calling the students to be assessed, need to be written. Second, the step-by-step process of administering the exercise needs to be written. This may begin with setting up any apparatus or materials needed in the assessment, correctly positioning the student(s) for the exercise, the actual material to be read to students (shown in a different manner than the directions to the assessment administrator), the probes to be used should students give no response or one that is not complete, how to move from one exercise to the next, and so forth. The trick is to "think through" the administration process one step at a time, continually asking yourself "what if?" types of questions. Once the directions are written, it may be helpful to administer them to someone else or have someone else administer the exercises to you.

Developing standard assessment administration directions that are complete and accurate is usually the result of trying the exercise out one or more times and noting areas of student confusion, responses that students provided which are vague or incomplete (and devising appropriate probes in order to elicit additional information from the student), and ways in which some or all students responded that were not anticipated. While some of this can be caught at the editing stage, tryouts of performance exercises are critical.

D. TRYING OUT THE ASSESSMENT EXERCISES

As mentioned above, trying out performance exercises is critical to the development of sound performance exercises. Unlike group-administered, paper-and-pencil exercises, where having sufficient numbers of students respond to each exercise in tryouts is critical, the most important part of trying out the performance exercise is to make sure that it is given to students by one or more well-trained individuals. In essence, it is important to train the assessment administrator as well as possible, certainly as well as will be done in the actual assessment. This means also that the tryout assessment administrator should be the same type of individual as will be used in the actual assessment, since a major purpose of the tryouts is to determine whether the exercise can be properly administered. It also means that draft assessment administration directions should be written and used for training the tryout assessment administrators, so that the effectiveness of the written procedures for training can be tried out and refined as well.

In addition, it is important to determine just how assessment administrator-specific the assessment administration directions are. This means that it may be more important to have ten to fifteen assessment administrators administer the performance exercise to five students each than to have five assessment administrators administer the exercise fifteen times each. While this will require extra work, both in terms of recruiting tryout assessment administrators and assessment administration, the payoff will be receiving a more accurate picture of the feasibility of the assessment administration process.

If it is feasible and if the resources permit, it is desirable to assess as many students as possible. If a sample of schools and students can be drawn and multiple assessment administrators used as suggested above, having a sample of as few as 300 students at the tryout stage may be sufficient. However, larger sample sizes are always desirable, particularly to assure that the range and types of responses to the exercise will represent the range and type of responses to be collected in the actual

assessment, as well as to examine the differences in performance among groups of students. The more complex and varied the types of responses anticipated, the larger and more varied the tryout sample should be.

E. DEVELOPING SCORING GUIDES

The development of scoring guides for performance exercises is similar in some respects to the development of scoring guides for any open-end exercise. However, the types of responses may vary widely. Responses to performance exercises may range from psychomotor responses that are observed and rated live (or on video tape), recorded on audiotape, observed, recorded by the assessment administrator or written by the student being assessed. The first essential element to developing sound scoring guides is having obtained an adequate number and range of responses in the tryout phase of the project. As mentioned above, the number and types of responses anticipated will help determine the sample size required to achieve this goal.

Once the student responses are gathered, someone needs to review the responses and attempt to score them according to the criteria and preliminary scoring guide developed by the exercise writer. This may be the exercise editor who does this step, or it may be another individual associated with the project. However, this step is generally best handled by one or two individuals. The persons should be trying to verify the preliminary scoring guide by trying to find samples of each type of anticipated response, as well as samples of student responses that do not fit the preliminary scoring guide.

Once the initial work on the scoring guide is complete, an expert panel of judges should be convened. This expert panel may be composed of classroom teachers in the area, curriculum specialists and subject-matter experts from the university level. A substantial amount of training may be necessary for the scoring panel to learn to reliably use the scoring guides developed for each exercise. This training should be documented so that it can be replicated when the subject area is re-assessed in the future.

This panel should be asked to review each exercise, confirm the preliminary judgments for each student response selected to illustrate each score scale point, and to discuss those student responses that did not appear to be scorable according to the preliminary scoring guide. The panel may note changes that need to be made in the assessment administration process, such as the addition of probes for particular student responses or places where some of the students being assessed appeared confused. Once completed, the expert panel review should confirm the

scoring guide or serve as the basis for changes in it. The result will be having an exercise that is feasible to administer and which can be scored appropriately.

F. STATISTICAL AND TECHNICAL ISSUES

Because the performance assessments (whether open-ended exercises or ones requiring more intricate, individual administration) require more time to administer, in a fixed period of time for assessment, fewer performance assessments than conventional, multiple-choice exercises can be administered during that time. This combined with the intended broad nature of many performance assessments raises important issues that need to be examined as the exercises are being tried out. As the assessment tryouts results are being analyzed, there are several technical issues that the developers should be concerned with. These include generalizability, bias, equating, and scaling. Each of these issues should be investigated during and after the tryouts.

Generalizability concerns whether any set of exercises chosen represent the total domain under investigation. Since a relatively few performance assessments are typically given, the tryouts should investigate how many exercises of the type(s) used are needed to obtain stable estimates of student achievement.

Bias is also of concern. There are several types of bias of concern. The first is whether there are differences in performance by various subgroups of students. During tryouts, it will be important to examine the differential item functioning (DIF) of the exercises by gender and racial-ethnic group. It may be important to examine the performance of the exercises by region and community-type. Both statistical DIF analysis and expert judgment should be used to review the exercises. One final note: as the exercises are being tried out, the issue of response bias should also be investigated. This refers to the differences among students with the *mode of response* used in the exercises. For example, does the written or performance response format disadvantage certain students who would be able to perform adequately on the exercise if presented in a different mode? This should also be investigated during the tryouts by interviewing students following the tryouts, as well as trying out the exercise with different response modes.

Scaling and equating are also issues that should be investigated following tryouts. Scaling is important because the exercises will undoubtedly be reported together as an overall level of performance, either the performance assessments alone or in conjunction with other assessments used. In addition, due to the unique, "memorable" nature of these exer-

cises, it will be important to use new exercises during each actual administration. Each of these will require that the set of performance assessments be scaled together, at least among the performance assessments (and with the other exercises to be used). By placing the exercises on the same scale, it will be possible to select subsets of the exercises for use each year, as well as to equate the different forms of the assessment.

G. REFINING THE EXERCISES AFTER TRYOUTS

Once tryouts are complete, the student samples have been scored and the scoring panel has made any suggestions for improvement of the exercise, the exercise editor can make any final changes in the exercise. Presumably, at this stage, the changes proposed are minor in nature. If they are not, as is sometimes the case, then the process given above really should be repeated. While there may be a temptation to assume that any major changes in the exercises have been made and will correct any deficiencies in the exercises, this is really not known until new student data is collected. After all, if it didn't work the first time, why are you certain that it will work this time?

The assumption is that it is far cheaper to try out the exercise one more time than to put a faulty exercise in the large-scale assessment program and not have the exercise work. At best, this will be a waste of money and effort; at worst, the exercise may work so poorly that it may disrupt students' responses to other exercises. Hence, major changes in the assessment administration directions, student tasks, stimulus materials or scoring guides suggest additional tryouts are needed before the exercise is used in the large-scale program.

III. Preparation for Assessment Administration

Once the performance exercises have been developed and tried out, the next important set of activities is getting ready for the actual assessment. This involves selecting the schools that will participate, preparing the schools for participation in the assessment and training the individuals who will gather the data from students. This is a stage that is often not given enough attention in planning and conducting performance assessments, yet is a vital one for the gathering of useful information about students.

A. DRAWING THE SAMPLES OF SCHOOLS AND STUDENTS

Since the assessment materials that have been developed and refined are intended for large-scale assessment use, it is presumed that the exercises will be given to some or all of the students at one or more grade or age levels. The original assessment plan will help to determine how and to whom the exercises will be administered. If each of the exercises is to be administered to all students at a grade or age level, no sampling is needed and the assessment designed can move on to the next stage. However, if all students at a particular grade or age level will not be assessed, of if they will not all take the same assessment, then some type of sampling procedure will be needed.

Sampling is an activity that most large-scale assessment programs have a contractor skilled in sampling design perform for them. It is a process that is relatively easy to carry out at the school level, particularly if the state or district already has some process of stratifying the state or district. If this is the case, then it should be relatively easy to draw a sample of schools. Given below is one example of how sampling at the school and student levels can be carried out. The actual procedures used, and the complexity of these, is highly dependent, however, on the design and purposes for the assessment. The example shown is for a low stakes assessment in which only samples of schools and students will participate and results will be reported only at the statewide level.

It is better to include as many schools as possible in the sample by keeping the number of students to be assessed in any one school relatively low. Using a figure of 20 students per school and an overall minimum figure of 500 students per assessment package will require a minimum of 25 schools per package. One way to economize on travel expenses is to administer more than one assessment package in a school, with 20 different students in each school taking each assessment package. Once this is decided, the number of schools to be selected can be determined. By dividing this number by the number of sampling strata, the number of schools to be selected from each sample stratum can be determined.

Spaced sampling is one easy procedure for selecting the actual schools and students to participate. Since there are cases where selected schools may not have any students of the eligible grade or age level, or the school may refuse to participate, it will be helpful if a set of alternate/replacement schools is selected for each stratum at the time that the original sample of schools is drawn.

At the student level, drawing a spaced sample of students is equally easy and can be handled by a trained assessment administrator. Starting with

a complete list of eligible students, the assessment administrator first determines the total number of eligible students. Then, the administrator divides this by the number of students to be assessed on a assessment package in the school. This results in a constant, which can be labeled *c*. Then, the assessment administrator picks a random number between *1* and *c*; this is the first student included in the assessment for that assessment package. Then, the assessment administrator counts down the list of students, selecting every *nth* student; these are the remaining students who will be assessed on the assessment package. Since some of these students may be absent or unavailable for assessment, the assessment administrator should also select some additional, alternate students. Since this process is so easy, the assessment administrator can select the sample of students on the day of assessment, presuming that a complete list of students (in any order) is available on the day of assessment.

B. PREPARING THE SCHOOLS FOR ASSESSMENT ADMINISTRATION

The preparation of the school for the assessment begins with the notification that the school has been selected for participation in the assessment. It will help the school if the notification letter includes details such as when the assessment will occur, what will be involved in the assessment, what the school's responsibilities will be, who will administer the assessment and so forth. The more specific this notification letter can be, the fewer questions will be raised and the easier it will be to receive school cooperation.

Schools should be told the approximate time when the assessment will occur, who will be contacting them to make additional arrangements, what process will be used to draw the sample of students (and why they need to have a complete listing of all eligible students on the day of the assessment) and what facilities and equipment they will need to provide (e.g., a quiet room, two chairs, a table and a VCR and television monitor).

C. TRAINING THE ASSESSMENT ADMINISTRATORS

A major key to the success of the entire performance assessment project is the quality of the individuals who are selected to conduct the assessment. If the resources are available, an organization that can provide such services will be contracted with and that organization will handle most of the details of selection of the assessment administrators and the training of them to administer each assessment package. Moreover, even in these cases, the following guidelines may be helpful to review to assure that all

important points have been considered. However, if the resources for a contract for assessment administration are not available, the sponsoring agency will need to directly carry out the following activities.

Locating suitable assessment administrators may be the most challenging aspect of carrying out a performance assessment without having a contractor to administer it. One technique that has worked is to hire persons who would normally substitute teach. Since the school districts that will be in the sample are known at this stage, it is possible to contact these districts and ask them to name one or two persons who would substitute in their district in the subject area(s) to be assessed. These individuals may be interested in the assessment activity and be willing to be trained to administer the assessment in that district (and, perhaps, in adjacent ones). Another technique that works is to use the colleges and universities in the state as the network from which to draw. In this case, each university is asked to name a team made up of one or more faculty members, plus one or more graduate students. This works particularly well in the cases where the state has a regional university system. In one state, the university curriculum teams not only carried out the assessment, they also were available to other local districts not in the state sample that wished either to be assessed or that wished to be trained in administering the tests. The administration of a performance assessment can serve as an excellent professional development activity so that some local districts will support their staff in learning to administer the performance tests. In other contexts, this might not be possible if possible assessment administrators demand substantial honoraria or union contracts do not permit volunteer work.

In any case, once the assessment administrators have been identified, training should occur. Depending on the nature of the assessment, as well as whether student responses are simply going to be recorded in some fashion for later scoring or will be scored on the spot during the assessment, the training may be rather extensive or relatively brief. Such training sessions have been as short as a day and as long as four days. In any case, the same material is reviewed. First, the assessment administrators are briefed on the project and the steps that have taken place prior to the training session. Second, they are informed how the state sample was drawn and how they are to draw the sample of students (see above). Next, they are given their assignment of schools and told who and how to contact the district assessment coordinator and the building assessment coordinator.

Fourth, each assessment exercise that they are to administer at each grade level it is given is reviewed. This is done first by literally reading

a copy of the assessment administration guide to them. Then, each exercise is demonstrated to them and finally, they have a chance to practice administering each exercise to one another. They are also reminded to re-review and practice the assessment administration prior to administering the assessment the first time. This practice is particularly important where the assessment administrator must coordinate the use of manipulatives or equipment during the assessment. If possible, plenty of time is allowed for this practice assessment administration. Where time permits, arrangements are made to have either adults or children available in order to give the assessment administrators simulated practice in assessment administration.

Once the assessment administrators are comfortable with the administration of the tests, and the manipulatives and equipment that are used in them, the focus of the assessment administrator training shifts to the recording and scoring of student responses. In some areas, this may be relatively minor. For example, in an assessment of music, if student responses are tape-recorded, this step probably can be omitted, since the assessment administrators have already practiced using tape recorders as part of learning to administer the exercises. In the case of a physical fitness assessment where the assessment administrator must score student responses as the student responds, however, training may take one or more days.

In the case of physical fitness, where the assessment administrator needs to learn to score about twenty different exercises, two techniques were used. First, various student responses were videotaped, both at regular and slow motion speed. This allowed the development of both a training tape and a validation tape for each exercise. The training tape showed regular and slow motion performances for each score scale point. Second, accompanying the training videotape is a training manual that uses stick figures to illustrate each criterion for an acceptable or unacceptable performance. For example, there are several criteria for throwing a ball, including cocking the arm, the elbow leading the hand, breaking the wrist and the follow through. Each of these is illustrated in the training manual, and the assessment administrator saw examples of both good and poor performances on the videotape. The written criteria were reviewed extensively with the assessment administrator. Then, they practiced scoring other samples of student performance from the training videotape for each exercise.

Once the assessment administrators had been trained on each exercise separately and were able to properly score each of the samples on the training videotape, they were given the validation videotape to score.

This tape presented several student responses in the order in which they occurred in an assessment package. Since an assessment package could contain several different exercises in a row and the assessment administrator would need to remember the different scoring criteria for each, the validation videotape presented the samples for each student together, just as they would appear to the assessment administrator as they administer the exercises live. Assessment administrators with acceptable scoring prowess were certified as assessment administrators; others were either cycled back through more assessment administration training or were dismissed from the assessment project. Fortunately, almost all of the assessment administrators were able to demonstrate their ability to score reliably the first time.

At this point, then, the assessment administration training was complete, and the assessment administrators were told to contact the district assessment coordinator to begin the process of making arrangements for the assessment administration.

IV. Assessment Administration

It may seem that a lot of work has already gone into the development and training for the performance assessment. And yet, the next phase is also a crucial one in assuring the feasibility and accuracy of performance assessment in the schools. The manner in which schools are contacted, students selected and assessed, and the assessment process is reviewed for improvement in the future can determine the success or failure of the entire process. Therefore, this is not a time to relax nor not pay attention to what the assessment administrators or schools are doing.

A. NOTIFICATION OF SCHOOLS ABOUT THE ASSESSMENT

The notification to the schools that were selected to take part in the assessment usually can take the form of a letter to both the district and the school coordinator. Be sure to mention in this letter any conditions needed for the assessment administration, such as equipment, facilities and so forth. This will allow the school to locate or secure the needed equipment or facility.

Once the assessment administrator has been trained, the assessment administrator should contact the district assessment coordinator, remind the district assessment coordinator that one or more schools were selected to take part in the performance assessment, indicate just

what the assessment will consist of and determine who the school assessment coordinator is. If some time has elapsed from the time that the initial letter to the district and school was sent, it may be helpful to send a reminder note to both shortly before the assessment administrator is to contact the district.

Once the district contact has been made (and it is critical that this occur before the building contact is made), the assessment administrator can contact the building assessment coordinator. Hopefully, this individual is aware that the performance assessment will be occurring in his/her building. The purpose of this contact is to schedule the date(s) and times for the assessment administration to occur, to remind the school coordinator to have a complete listing of students at the appropriate grade level(s) available on assessment day and what facilities and equipment will be needed. These contacts should be made before any assessment administration starts.

Any problems that are uncovered should be relayed to the state contact person. This might require the selection of a replacement school or for the state contact person to contact the district or school in order to secure cooperation and so forth.

B. MONITORING THE ASSESSMENT

Once all contacts have been made, the assessment administrators should begin the process of assessment administration. This will start with the assessment administrator drawing the sample of students to be assessed and a list of alternates. About an hour should be allowed for this to take place. Once the student list has been compiled, it may be helpful to have the school assign an aide or a student who can work with the assessment administrator to locate the students when needed for the assessment and bring them to the assessment administration site. A person who is familiar with the teachers, the class schedule and the physical layout of the school will be the most helpful.

This assessment administrator aide can bring students to the assessment administration site as needed and can help assure that each student is returned to class without undue disruption. The next student can be brought to the room just before the previous student is finished so that the assessment administration can flow quickly and efficiently from one student to another.

As the assessment administration process is taking place in the schools, it will be most helpful if the state contact person and others associated with the project but not involved directly in the assessment administra-

tion would select some schools in which to observe one or more students taking part in the performance assessment. Not only will the observers have an opportunity to observe assessment administration taking place, they can also discuss the assessment with some of the students following the assessment. This can provide invaluable insight on why and how students responded and their motivation and interest in the assessment, as well as ways in which the assessment administration and assessment administrator training processes can be improved in the future.

C. EVALUATING THE ASSESSMENT PROCESS

The monitoring process outlined above will be helpful in the evaluation of the performance assessment administration process. It may also be helpful to solicit comments directly from the assessment administrators, school coordinators and district coordinators. This evaluation process is important in order to determine whether there were any factors that affected student performance and, therefore, ought to be considered when interpreting the data that result from the assessment. It also can be helpful for the future when other assessments are being designed in order that problems encountered in this assessment are not repeated (or, at least, are better anticipated) in the next assessment.

V. Post-Assessment Administration Activities

Once the assessments of students on the performance exercises is completed, there still may be considerable work ahead in order to score, report and interpret the assessment information. If the performance assessment was scored while being administered, some of these steps can be omitted. Otherwise, these steps will be essential for compiling useful information.

A. TRAINING THE SCORERS OF OPEN-END EXERCISES

The first step in the scoring phase is to locate persons who can be trained to be scorers. In some cases, it is desirable to select and train actual classroom teachers in the scoring routines and processes. Such scorer training can be an excellent professional development activity. If this is done at a time when school is in session, then the number of days available from any scorer may be limited and it may be necessary to train more scorers. Also, it may be advantageous to conduct the scoring at a regional level rather than at a central location, so that the necessity for overnight accommodations is limited. This will allow a teacher, for

example, to drive to the scoring center each day and return home in the evening. One way in which regional scoring centers can be "housed" is on college campuses.

The next step is the development of the training package. This may begin with the development of a training package comparable to what was developed to train the assessment administrators who would have scored student responses in the field. This all takes place before the scorers are convened for the first time. First, various student responses are selected by an expert in the area to represent the various types of responses that students may have given (in the case of "holistic" scoring) or that illustrate each type of response (in the case of "primary trait" scoring). This may also lead to refinements in the scoring guide and rubrics.

Such expert judgments are next confirmed by an expert panel of judges; some of the samples will appear in the scoring guide pre-scored and will be used to train the scorers. The others (that are also pre-scored) will be used to judge the accuracy (reliability) of the scorers following the initial training. Depending on the nature of the assessment packages and how student responses were collected, it may be desirable to have the scorers score just one exercise at a time, or to be able to score all of the exercises in an assessment package for students one at a time. This will help to dictate the structure of the training sessions and the scoring process.

Once the scorers had demonstrated their ability to reliably score the exercise(s), training was complete. Scorers are then ready to begin the scoring process.

B. CONDUCTING THE SCORING OF OPEN-END EXERCISES

The scoring process requires a number of things to be pre-arranged for the scoring to flow smoothly. First, a determination would have to be made about whether there will be one or two scorers for each response. If more then one scorer will be used, what process will be used to resolve differences? How will the first scorer's scores be kept from the second scorer? Second, arrangements will need to be made to efficiently distribute booklets and/or other materials to be scored, plus rating sheets, to each of the scorers. It will undoubtedly be desirable to have runners available to distribute and collect materials. In addition, it is critical that persons who are well-organized be used to keep track of each booklet and/or other material to be scored to make certain that it has received all needed ratings.

Third, routine reliability checks should be built into the scoring process. There are a couple which are most important. Each scorer should be given a few of the pre-scored exercises without notice to score during the actual scoring session. If their ratings have "drifted" from those of the expert panel, it may be necessary for on-the-spot review of the scoring guides. It is also important for each scorer, later on in the scoring, to re-score a few responses that they scored earlier to make certain that scoring subsequent student responses has not caused their scoring standards to change. Again, it may be necessary to conduct an on-the-spot review of the scoring criteria.

Once the scoring design is determined, then the scoring training and actual scoring can commence. This may be completed in one session or subsequent sessions may be required. If more than one session is needed, subsequent review of the scoring standards may also be needed.

C. SUMMARIZING THE ASSESSMENT RESULTS

Once the students' responses to each of the performance exercises have been scored, the various scores need to be summarized and prepared for reporting. The summaries will be most efficient if the person(s) who will be doing the reporting and those persons directing the scoring discuss the needed and desired data summarization process before the scoring is conducted. This will allow the most efficient use of the scorers and the scoring contractors in preparing the data in the format(s) needed to summarize the data in ways that will lead most efficiently to the types of reports to be issued. Careful thought about this in advance can save hours of re-coding or re-entry of data.

D. REPORTING THE ASSESSMENT RESULTS

There are probably at least as many ways of reporting on the performance of students on performance exercises as there are number of exercises. However, one thing to keep in mind is that the typical reader of a report will probably be unfamiliar with the performance exercise, the subject area, the scoring criteria and so forth. What the typical reader is probably interested in is whether or not many students were able to correctly perform the required task, and if not, what were the typical performances that student gave. Explanations for low or unexpected performance are also usually of interest. If the reports of results, at least for the public and for classroom teachers, focus on these points, the report of results will be of interest to the widest audience.

There will be others who are interested in more detail about how students performed on the exercises and for these groups, reports of a more

complete and/or technical nature will be most beneficial. Others may be interested in obtaining the actual samples of student responses in order to conduct secondary analyses of the data. With proper safeguards for student and teacher privacy, these data can be made available, and subsequent reports that these researchers make can also be publicized.

E. INTERPRETING THE ASSESSMENT RESULTS

The reports alluded to above suggest that more than mere numbers will be provided. In addition, most audiences want to know not only how well students performed, but also why students performed as they did, were the experts surprised in any way by the level or types of student performances, and perhaps most importantly, what do the experts think needs to be done to help students improve. These important questions should be covered in the interpretation of the performance assessment results.

It may also be helpful to integrate the interpretation of both the performance and non-performance exercises into a single interpretative report. This could allow, for example, knowledge questions from a multiple-choice assessment to be reported in the context of whether or not students were able to apply that knowledge in various applications assessed in the performance exercises. This may lead to clearer interpretations of the results without diminishing the importance of the performance measures. Just as in reporting, keep in mind that different depths of interpretation may be desired and needed by different audiences, so that multiple interpretive reports may be best for meeting different information needs.

Summary

This guide to performance assessments has illustrated the steps in creating, trying out, administering, scoring, and reporting performance assessments for large-scale assessment use. While such assessments are not easy to develop, administer, score and interpret, much is known about the steps that need to be carried out, and many different organizations at the national, state and local levels have already conducted successful assessments in a wide variety of subject areas. These assessments range from the simple to the complex and from ones that cost little (a few thousand dollars) to a considerable amount (hundreds of thousands to millions of dollars).

While performance assessment may be new to some people (news media stories have characterized how it has been recently "discovered"

by some), it is not new nor is it untried. National and statewide performance assessments were successfully conducted in a reliable and cost-efficient manner decades ago. As this guide has hopefully illustrated, performance assessment is feasible and manageable. Such assessments are vitally needed in the assessment landscape so that those interested in assessing what students are capable of doing have access to more complete information on student performance. While the steps are more complex and more involved, such assessments are important in the determination of what skills our students need to have and whether or not they do in fact have them. Performance assessment is an important adjunct to overall large-scale assessment strategies.

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE
COUNCIL OF CHIEF STATE
SCHOOL OFFICERS

INTRODUCTION

IMPLEMENTATION MODELS AND PROTOTYPES

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

POLICY

7

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

BIBLIOGRAPHY

References

American Association for the Advancement of Science. (1996). *Science Education Reform for All (SERA): A Look At How State Departments of Education Are Infusing Equity and Excellence Into PreK-12 Systemic Reform.* Washington, DC: Author.

American Federation of Teachers. (1997). *Making Standards Matter 1997: An Annual Fifty-State report on Efforts to Raise Academic Standards.* Washington, DC.

American Federation of Teachers. (1997). *What Students Abroad Are Expected to Know About Mathematics.* Washington, DC: Author.

American Federation of Teachers. (Spring 1996). *American Educator: Spring Brings New Life to the Standards Movement and a Historic Opportunity to Strengthen Our Schools.* Washington, DC: Author.

Aspen Systems Corporation, National Assessment Governing Board & The National Center for Education Statistics. (1995, October). *Joint Conference on Standard Setting for Large-Scale Assessments: Executive Summary, Volume I.* Washington, DC: Authors.

Council of Chief State School Officers. (1997). *Mathematics and Science Content Standards and Curriculum Frameworks: State Progress on Development and Implementation.* Washington, DC: Author.

Council of Chief State School Officers. (1996, October). *Key State Education Policies on K-12 Education: Content Standards, Graduation, Teacher Licensure, Time and Attendance.* Washington, DC: Author.

Council of Chief State School Officers. (1996). *Key State Education Policies on K-12 Education: 1996, Appendix A. State Student Assessment Programs in Mathematics, Language Arts, and Science.* Washington, DC: Author.

Council of Chief State School Officers. (1995, May). *State Curriculum Frameworks in Mathematics and Science: How Are They Changing Across the States?* Washington, DC: Author.

Education Week, Pew Charitable Trusts. (January 22, 1997). *Quality Counts: A Report Card on the Condition of Public Education in the 50 States.* Washington, DC: Authors.

Massell, D., Kirst, M., & Hope, M. (1997, March). *Persistence and Change: Standards-Based Systemic Reform in Nine States.* Consortium for Policy Research in Education, RB-21.

McLaughlin, M.W., Shepard, L.A., & O'Day, J. (1995). *Improving Education Through Standards-Based Reform.* Stanford, CA: National Academy of Education.

National Association of State Boards of Education. (1997). *The Full Measure: Report of the NASBE Study Group on Statewide Assessment Systems.* Alexandria, VA: Author.

National Association of State Boards of Education. (1996, February). *Issues in Brief: The Fall and Rise of Standards-Based Education.* Alexandria, VA: Author.

National Council of Teachers of Mathematics. (1997). *Improving Student Learning in Mathematics and Science: The Role of National Standards in State Policy.* Washington, DC: Author.

National Governors Association. (1996). *Setting Educational Standards: Experiences in Four States.* Washington, DC: Author.

Pritchard, I. (Summer 1996). Judging Standards in Standards-Based Reform. *Perspective*, 8(1).

Tool Kit

Evaluating the Development and Implementation of Standards

DEVELOPED BY THE
COUNCIL OF CHIEF STATE
SCHOOL OFFICERS

INTRODUCTION

IMPLEMENTATION MODELS AND PROTOTYPES

ANALYSIS OF STATE CONTENT STANDARDS

CRITERIA FOR STANDARDS

POLICY

8

ASSESSMENT GUIDELINES AND RESOURCES

BIBLIOGRAPHY

NOTES

306